

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

COMPLEXITÉ PALINDROMIQUE DES MOTS ET DES ARBRES

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN MATHÉMATIQUES

PAR

NADIA LAFRENIÈRE

JANVIER 2016



## REMERCIEMENTS

J'aimerais d'abord remercier mes deux directeurs, Srečko Brlek et Xavier Provençal pour leur présence, leur support et l'opportunité d'aller en France. Srečko, tu m'as fait aimer la combinatoire des mots. Ton support, moral et financier, ta légèreté et ton amour de la bouffe sont des qualités grandement appréciées. Merci Xavier pour l'accueil à Chambéry, ta minutie, ta disponibilité, l'introduction à la géométrie discrète et les excursions à la montagne.

Merci à Christophe Reutenauer et Alexandre Blondin Massé qui ont accepté de lire mon mémoire et qui ont formulé des remarques visant son amélioration. Je leur en suis très reconnaissante.

Sébastien Labbé m'a outillée pour mieux utiliser un ordinateur et m'a introduite au développement de Sage, en plus de s'être intéressé à ma recherche; je l'en remercie grandement!

Je ne saurais assez remercier celles et ceux avec qui j'ai organisé de multiples projets : Sébastien Ouimet, Mélodie Lapointe, Rachel Beaulieu-Salamido, Alex Provost, Jean-François Arbour et Stéphanie Schanck. Vous êtes de formidables amies et amis, dont j'ai toujours à apprendre de l'ouverture. Merci d'avoir partagé mes réflexions tout au long de mon cheminement en mathématiques et de m'accompagner dans mes idées.

Je remercie Franco Saliola et Alexandre Blondin Massé pour leur bonne humeur

et leur enthousiasme contagieux. De par votre enseignement ou les activités de recherche, vous avez su faire grandir ma passion pour les mathématiques et la combinatoire.

Merci également à Christophe Hohlweg pour ses encouragements tout au long de mes études. Tu es le premier à avoir cru en moi à un moment où je n'aurais pas envisagé poursuivre mes études aussi loin. Merci à Jérôme Tremblay et Johanne Patoine de nous épauler dans chaque problème qui survient dans une vie de laboratoire et d'en faire un endroit aussi agréable.

Herman Goulet-Ouellet, Émile Nadeau et Pauline Hubert font du LaCIM un endroit où il fait si bon travailler. Merci! Merci à celles et ceux qui ont fait de mon séjour au LAMA un meilleur moment : Rodrigo Dorantes-Gilardi, Charlotte Perrin, Michel Raibaut et Tom Hirschowitz.

Merci enfin à mon père, Gilbert, et ma sœur, Laurence, pour leurs encouragements sans limites.

## TABLE DES MATIÈRES

TABLE DES FIGURES . . . . .	vii
LISTE DES TABLEAUX . . . . .	ix
RÉSUMÉ . . . . .	xi
ENGLISH ABSTRACT . . . . .	xiii
INTRODUCTION . . . . .	1
PRÉLIMINAIRES . . . . .	5
CHAPITRE I	
COMPLEXITÉ ET DÉFAUT PALINDROMIQUES DES MOTS INFINIS.	19
1.1 Nombre de palindromes et défaut . . . . .	19
1.2 $\sigma$ -Défaut des mots finis . . . . .	22
1.2.1 Un algorithme pour calculer le $\sigma$ -défaut . . . . .	24
1.3 $\sigma$ -Défaut des mots infinis . . . . .	26
1.3.1 $\sigma$ -Défaut des mots périodiques . . . . .	27
1.3.2 $\sigma$ -Défaut des mots apériodiques . . . . .	35
CHAPITRE II	
UNE IDENTITÉ DE BRLEK ET REUTENAUER. . . . .	39
2.1 Complexité palindromique et complexité en facteurs . . . . .	39
2.2 L'identité en question . . . . .	41
2.3 Mots finis . . . . .	42
2.4 Mots périodiques . . . . .	44
2.5 Mots dont le langage est fermé par $\theta$ . . . . .	48
2.5.1 Exemples de mots dans cette classe . . . . .	50
2.5.2 Théorème principal . . . . .	52
2.6 Mots uniformément récurrents . . . . .	69
2.7 Points fixes de morphismes primitifs . . . . .	71
2.8 Une identité vraie pour les mots récurrents? . . . . .	72
2.8.1 Mots ayant un nombre fini de palindromes . . . . .	72
2.8.2 Un problème ouvert . . . . .	73
CHAPITRE III	

LA COMPLEXITÉ PALINDROMIQUE DES ARBRES. . . . .	75
3.1 Une famille d'arbres avec beaucoup de palindromes . . . . .	78
3.1.1 Quelques éléments de combinatoire additive . . . . .	79
3.1.2 La construction en « forme de peigne » . . . . .	82
3.2 Vers une borne maximale . . . . .	85
3.2.1 Arbres dans $\mathcal{T}_1$ et $\mathcal{T}_2$ . . . . .	86
3.2.2 Arbres dans $\mathcal{T}_3$ et $\mathcal{T}_4$ . . . . .	87
3.2.3 Hypothèses pour la construction d'arbres avec un grand nombre de palindromes distincts . . . . .	95
3.2.4 La résolution du problème général . . . . .	98
CONCLUSION . . . . .	101
BIBLIOGRAPHIE . . . . .	103

## TABLE DES FIGURES

Figure	Page
0.1 Un graphe simple, avec (à gauche) et sans (au centre) cycle. À droite, un graphe étiqueté par l'alphabet $\{a, b, c\}$ . . . . .	14
0.2 Graphe de Rauzy $\Gamma_3(\text{Fib})$ et graphe de Rauzy réduit $\Gamma'_3(\text{Fib})$ du mot de Fibonacci. . . . .	16
1.1 Le mot de Christoffel qui code la droite de pente $\frac{5}{8}$ . . . . .	20
1.2 L'ajout d'une lettre à un mot crée au plus un $\sigma$ -palindrome original.	23
1.3 Les mots <i>aquaplane</i> et <i>ombrelle</i> vus comme mots circulaires. . . . .	29
1.4 Le chevauchement de $\sigma$ -palindromes implique une structure particulière. . . . .	33
1.5 Le produit de deux $\sigma$ -palindromes <i>appétit</i> et sa représentation comme mot circulaire. . . . .	34
2.1 Graphes $G_{\text{Id}}(\text{Fib}, 3)$ et $G_E(\text{Fib}, 3)$ du mot de Fibonacci. . . . .	59
2.2 Pour un facteur de longueur $n$ et une arête du graphe $G_\sigma(\mathbf{w}, n)$ donnés, il existe un unique facteur de $\mathbf{w}$ correspondant à l'étiquette.	60
2.3 Une chaîne de longueur $m + 1$ dans $G_\sigma(\mathbf{w}, n)$ . . . . .	61
2.4 Récurrence sur les facteurs spéciaux droits du mot $v$ . . . . .	64
2.5 Préfixe de $\mathbf{w}$ terminant par le premier mot de $\theta$ -retour complet de $u$ .	66
2.6 Structure de $r$ , un mot de $\theta$ -retour complet de $u$ dans $\mathbf{w}$ . . . . .	67
2.7 Mots de $\theta$ -retour complet de $s$ et de $\hat{s}$ dans $\mathbf{w}$ . . . . .	68

3.1	Un arbre filiforme représente une paire formée d'un mot et de son image miroir. . . . .	75
3.2	Le domaine fondamental du graphe de Christoffel associé au vecteur $(2, 3, 5)$ . . . . .	77
3.3	Un arbre avec 6 arêtes et 7 palindromes non-vides. . . . .	77
3.4	Le peigne $\mathcal{Q}_{p,\beta}$ . . . . .	82
3.5	L'arbre $\mathcal{Q}_p$ . . . . .	85
3.6	Un arbre de $\mathcal{T}_2$ . . . . .	87
3.7	La connexité de l'arbre $S$ est garantie par le fait que c'est une restriction d'un arbre de $\mathcal{T}_4$ . . . . .	89
3.8	La destruction d'une rupture sur la lettre $b$ . . . . .	92
3.9	La forme de l'arbre $S'$ . . . . .	93
3.10	La structure de l'arbre $T'$ , dont les chaînes maximales sont des palindromes. . . . .	96
3.11	Un triplet de nœuds dont les chaînes entre chaque paire forme un palindrome est isomorphe à une partie du peigne. . . . .	97
3.12	Un arbre double qui détermine le langage d'un autre arbre. . . . .	99



## LISTE DES TABLEAUX

Tableau		Page
0.1	Quelques mots de retour complet de facteurs du mot <i>mentalement</i> .	13
1.1	Lacunes et longueurs des plus longs $\sigma$ -palindromes suffixes des préfixes d' <i>institution</i> . . . . .	23
1.2	Lacunes et longueurs des plus longs $\sigma$ -palindromes suffixes des préfixes d' <i>appétit</i> . . . . .	34
2.1	Complexités palindromique et en facteurs du mot <i>institution</i> . . . . .	44
3.1	La table d'addition d'un ensemble de Sidon. . . . .	79



## RÉSUMÉ

Ce mémoire concerne l'étude des palindromes, soit les mots qui se lisent de la même façon de gauche à droite et de droite à gauche. On s'intéresse au nombre de facteurs palindromes d'un mot, c'est-à-dire de suites contiguës de lettres qui sont des palindromes. On élargit la portée de théorèmes connus aux  $\sigma$ -palindromes, une variante qui permet de considérer des transformations différentes de l'image miroir. On choisit alors  $\sigma$  comme une permutation involutive des lettres de l'alphabet et un  $\sigma$ -palindrome est un point fixe de la composition du miroir et de  $\sigma$ . On expose des algorithmes permettant de calculer le nombre de  $\sigma$ -palindromes et le  $\sigma$ -défaut, une mesure de densité en  $\sigma$ -palindromes originaux. On établit aussi une relation entre le  $\sigma$ -défaut et le nombre de facteurs et de  $\sigma$ -palindromes de chaque longueur, qui s'applique tant aux mots finis qu'aux mots infinis qui sont aussi périodiques, fermés par la composition de  $\sigma$  et de l'opérateur miroir, uniformément récurrents ou points fixes de morphismes primitifs.

On considère ensuite des arbres finis dont les arêtes sont étiquetées par des lettres. La trace d'un chemin de l'arbre définit un mot qui appartient au langage de l'arbre. On se penche sur le nombre maximal de palindromes que peut contenir un tel arbre. On remarque alors que ce nombre est plus grand que l'analogue pour les mots. On fournit enfin une borne asymptotique du nombre de palindromes dans les arbres ayant une certaine structure et on survole la résolution du problème général, récemment découverte.

**Mots-clefs :** Combinatoire des mots, palindromes, complexité palindromique, pseudopalindromes, défaut palindromique, arbres.



## ENGLISH ABSTRACT

**English title:** Palindromic Complexity Of Words And Trees

This thesis concerns the study of palindromes (words that read the same from left to right and from right to left). We are interested in the number of contiguous subsequences of letters that are palindromes. We extend some already known theorems to  $\sigma$ -palindromes, a variation that allows considering other transformations than the reversal. We choose  $\sigma$  to be an involutive permutation of the letters of the alphabet and we say that a  $\sigma$ -palindrome is a fixed point of the composition of the reversal and  $\sigma$ . We expose algorithms to compute the number of  $\sigma$ -palindromes and the  $\sigma$ -defect, a measure of density of original  $\sigma$ -palindromes. We establish a relation between the  $\sigma$ -defect and both the factor and palindromic complexities. This equation is true for finite words, as well as periodic words, uniformly recurrent words, fixed points of primitive morphisms and words that are closed under the composition of  $\sigma$  and the reversal.

We also consider finite trees that are labelled by letters. The trace of a path defines a word that belongs to the language of the tree. We study the maximum number of palindromes in such trees. We notice that this number is larger than the maximum number of palindromes in words. We provide, finally, an asymptotic bound for the number of palindromes in trees having a given structure, and explain how the global problem was recently solved.

**Keywords:** Combinatorics on words, palindromes, palindromic complexity, pseudo-palindromes, palindromic defect, trees.



## INTRODUCTION

La question de l'étude des régularités dans les suites de symboles a été posée formellement pour une des premières fois par Axel Thue (Thue, 1906; Thue, 1912). Il cherchait alors à connaître l'existence de mots sans carré (des répétitions d'un même mot, comme *couscous*). Ses travaux ont non seulement contribué à l'étude du sujet, mais Axel Thue est surtout reconnu pour avoir introduit une étude systématique des mots.

Si les carrés ont été étudiés, ce ne sont pas les seules répétitions qui ont été à l'honneur de travaux effectués surtout depuis la deuxième moitié du XX<sup>e</sup> siècle. Les chevauchements, les cubes et les plus grandes puissances ont également fait l'objet de recherches. Certains motifs sont aussi étudiés à travers la périodicité qu'ils induisent : un résultat de Nathan J. Fine et Herbert S. Wilf (Fine et Wilf, 1965) permet en effet de résoudre des équations sur les occurrences multiples de motifs.

Bien entendu, une question qui s'est posée est de savoir si un motif existe toujours – c'était d'ailleurs la question d'Axel Thue – mais aussi de compter ses occurrences.

D'autres formes de répétitions ont été étudiées, parfois à permutation des lettres près. Un motif intéressant est le palindrome, qui est un mot ayant un axe de

symétrie en son centre. Il est d'ailleurs l'objet de ce mémoire, de pair avec une de ses généralisations, les pseudopalindromes ou  $\sigma$ -palindromes.

L'étude des complexités palindromique et en facteurs permet aussi d'obtenir des résultats intéressants en théorie des nombres. Par exemple, pour tout nombre algébrique irrationnel, la complexité en facteurs – le nombre de facteurs de chaque longueur – des décimales du nombre est bien plus que linéaire (Adamczewski *et al.*, 2004) et il est conjecturé que tous les nombres algébriques irrationnels atteignent la complexité maximale (Adamczewski et Bugeaud, 2010).

Les palindromes jouent aussi leur rôle. Par exemple, soit un mot infini  $x$  sur un alphabet constitué d'entiers positifs et dont les lettres représentent l'expansion en fraction continue :

$$[x] = \frac{1}{x_1 + \frac{1}{x_2 + \frac{1}{x_3 + \dots}}}$$

Si  $x$  a des préfixes palindromes arbitrairement longs, alors le nombre  $[x]$  est soit quadratique ou transcendant (Adamczewski et Bugeaud, 2007). D'autres résultats sur les occurrences de palindromes en théorie des nombres peuvent être trouvés dans (Allouche et Shallit, 2000), notamment.

Les  $\sigma$ -palindromes – une généralisation des palindromes, où on remplace l'opérateur miroir par n'importe quel antimorphisme involutif – ont été introduits en combinatoire des mots dans le contexte de l'étude de la périodicité (Anne *et al.*, 2005; de Luca et de Luca, 2006). Parallèlement, ils ont été largement étudiés en bio-informatique : appelés *palindromes de Watson-Crick* (Kari *et al.*, 2002; Husini *et al.*, 2003; Kari et Mahalingam, 2010), ils représentent, dans le langage de la combinatoire des mots, une paire de mots qui sont l'image l'un de l'autre par la composition du miroir et d'une involution sur les lettres. L'alphabet  $\{A, C, T, G\}$



code les suites de nucléotides des brins d'ADN. Dans la structure à double hélice antiparallèle de l'ADN, les nucléotides de cytosine (C) et de guanine (G) interagissent entre eux à travers trois liaisons d'hydrogène, alors que ceux d'adénine (A) et de thymine (T) se combinent à travers deux liaisons d'hydrogène. Ainsi, pour  $\sigma : \{A \leftrightarrow T, C \leftrightarrow G\}$ , deux brins d'ADN s'apparient complémentaiement; les nucléotides du deuxième sont l'image par l'antimorphisme involutif  $\rho \circ \sigma$  de ceux du premier, où  $\rho$  symbolise le miroir.

Ce mémoire est divisé de la façon suivante. La terminologie et des propositions célèbres en combinatoire des mots sont d'abord présentées. Au chapitre 1, on étudie le nombre de  $\sigma$ -palindromes dans les mots finis et infinis et on montre comment le calculer. La notion de  $\sigma$ -défaut est aussi introduite comme une mesure de densité de  $\sigma$ -palindromes distincts, particulièrement dans les mots infinis. Une façon explicite de le calculer est élaborée et on présente quelques résultats sur les  $\sigma$ -palindromes et le  $\sigma$ -défaut des mots périodiques, sturmiens et le mot de Thue-Morse. Une partie du contenu de ce chapitre a été publiée dans la revue *Fundamenta Informaticae* (Brlék et Lafrenière, 2014).

Le chapitre 2 s'attarde au lien entre les complexités en facteurs et  $\sigma$ -palindromique de certaines familles de mots. On lie ces deux valeurs avec le  $\sigma$ -défaut, défini au chapitre 1. Ce faisant, on généralise une identité de Srečko Brlek et Christophe Reutenauer d'abord définie pour les palindromes traditionnels, en établissant l'exactitude d'une relation pour les mots finis, périodiques, uniformément récurrents, points fixes de morphismes primitifs et fermés par un antimorphisme involutif.

Le chapitre 3, finalement, introduit le langage d'un arbre et compte le nombre maximal de palindromes qu'il contient. Il présente des résultats asymptotiques, que nous avons publiés récemment (Brlek *et al.*, 2015), sur le décompte des palindromes. On présente de plus une conjecture sur le nombre maximal de palindromes dans n'importe quel arbre, et sa résolution publiée par une autre équipe de chercheurs (Gawrychowski *et al.*, 2015).

## PRÉLIMINAIRES

Un *alphabet* fini  $\Sigma$  est un ensemble fini de symboles. Ses éléments sont appelés *lettres* et un *mot*  $w = w_1w_2 \dots w_n$  est une suite finie de lettres de  $\Sigma$ .

L'ensemble de tous les mots est le monoïde libre sur  $\Sigma$  et est dénoté  $\Sigma^*$ . L'élément neutre du monoïde est noté  $\varepsilon$  et désigne le *mot vide*, et l'opération associative est la concaténation. La *longueur d'un mot*  $w$ , notée  $|w|$ , est son nombre de lettres.

On définit le demi-groupe des mots non-vides  $\Sigma^+$  comme l'ensemble des mots finis à l'exception de  $\varepsilon$ .

De façon analogue, on définit un *mot infini* comme une suite ordonnée infinie de symboles de l'alphabet. On note l'ensemble des mots infinis  $\Sigma^\omega$ .

**Facteurs.** Un *facteur* d'un mot  $w$  est une sous-suite contiguë de lettres de  $w$ . Un facteur de  $w$  qui apparaît au début du mot est appelé *préfixe*, et un qui est placé à la fin du mot est un *suffixe* de  $w$ . On utilise la notation  $\mathcal{L}(w)$  pour désigner le langage de  $w$ , c'est-à-dire l'ensemble de ses facteurs. Sa restriction aux facteurs de longueur  $n$  est notée  $\mathcal{L}_n(w)$ .

La complexité en facteurs est la fonction

$$\begin{aligned} \mathcal{C} : (\Sigma^* \cup \Sigma^\omega) \times \mathbb{N} &\rightarrow \mathbb{N} \\ (w, n) &\mapsto \# \mathcal{L}_n(w). \end{aligned}$$

*Exemple 0.0.1.* L'ensemble des facteurs du mot *allô* est

$$\{\varepsilon, a, l, \hat{o}, al, ll, l\hat{o}, all, ll\hat{o}, all\hat{o}\}.$$

La fonction de complexité du mot vaut  $(1, 3, 3, 2, 1)$  pour le quintuplet d'entiers  $(0, 1, 2, 3, 4)$ , puis 0 pour toutes les autres valeurs.

**Proposition 0.0.2** (Lemme 7.1, (Morse et Hedlund, 1938)). *Pour tout mot infini  $w$ ,  $\mathcal{C}(w, n)$  est une fonction croissante.*

*Démonstration.* Pour tout facteur  $u$  de longueur  $n$ , il existe au moins un facteur de longueur  $n + 1$  qui débute par  $u$ , par exemple celui qui débute à la même position que la première occurrence de  $u$ .  $\square$

Un facteur  $u$  est dit *spécial gauche* s'il existe au moins deux lettres  $a$  et  $b$  telles que  $au \in \mathcal{L}(w)$  et  $bu \in \mathcal{L}(w)$ . On définit de manière similaire les *facteurs spéciaux droits*.

**Mots périodiques.** Une *période* d'un mot  $w$  est un entier  $m \leq |w|$  tel que  $w_i = w_{i+m}$  pour tout  $i \leq |w| - m$ . Un facteur  $u$  est dit *primitif* s'il n'est pas une *puissance* d'un autre mot, c'est-à-dire s'il n'existe pas d'entier  $k \geq 2$  et de mot  $v$  tels que  $u = v^k$ .

Si un mot  $w$  possède une période, il est dit *périodique*. Tous les mots finis sont périodiques (de période  $|w|$ ). Un mot  $w$  est dit *ultimement périodique* s'il est la concaténation d'un mot fini  $u$  et d'un mot périodique  $v^\omega$  :  $w = uv^\omega$ .

*Exemple 0.0.3.* Le mot *entente* a pour plus petite période l'entier 3 : *ent · ent · e*.

*Exemple 0.0.4.* L'écriture décimale de  $\frac{11}{28}$  est ultimement périodique :  $0.39(285714)^\omega$ . Par ailleurs, l'écriture décimale de tout nombre rationnel est ultimement périodique.

**Réurrence.** Pour un mot fini  $w$  et une lettre  $a$ , on note  $|w|_a$  le nombre d'occurrences de  $a$  dans  $w$ . On note de façon analogue le nombre d'occurrences du facteur  $u$  dans le mot  $w$  par  $|w|_u$ .

Un mot infini  $\mathbf{w}$  est dit *récurrent* si  $|\mathbf{w}|_u = \infty$  pour tout facteur  $u$ . Il est dit *uniformément récurrent* si de surcroît la distance entre deux occurrences consécutives de n'importe quel facteur est bornée. Par exemple, les mots périodiques sont uniformément récurrents.

**Proposition 0.0.5** ((Allouche et Shallit, 2003), proposition 10.8.4). *Un mot est récurrent si et seulement si chacun de ses préfixes a au moins deux occurrences.*

**Morphismes de mots.** Un morphisme de monoïdes est une fonction qui préserve la structure :

$$\begin{aligned} \psi : M &\rightarrow M' \\ uv &\mapsto \psi(u)\psi(v). \end{aligned}$$

Il est facile de voir que l'identité satisfait les conditions pour être un morphisme.

Pour le monoïde libre  $\Sigma^*$ , un morphisme de mots est une fonction  $\psi : \Sigma^* \rightarrow \Sigma^*$  qui a la propriété d'être compatible avec la concaténation :  $\psi(uv) = \psi(u)\psi(v)$ , pour tous  $u, v \in \Sigma^*$ .

Un morphisme de mot  $\psi$  est dit *primitif d'ordre  $k$*  si pour toute lettre  $\alpha$ ,  $\psi^k(\alpha)$  contient toutes les lettres de l'alphabet.

*Exemple 0.0.6.* Considérons l'alphabet  $\{a, b\}$  et soit le morphisme

$$\mu : a \mapsto ab, b \mapsto ba.$$

Ce morphisme est primitif d'ordre 1. En effet, les deux lettres de l'alphabet apparaissent dans  $\mu(a)$  et dans  $\mu(b)$ .

Ce morphisme sera présenté à plusieurs reprises dans ce mémoire, étant donné qu'il permet d'engendrer le mot de Thue-Morse  $\mathbf{t} = \mu^\infty(a)$  (suite A010060 de l'On-line Encyclopedia of Integer Sequences (OEIS Foundation Inc., 2011)). Il a été d'abord énoncé par Eugène Prouhet ((Prouhet, 1851), cité dans (Berstel et Perrin, 2007)), utilisé plus tard par Axel Thue (Thue, 1912), puis largement étudié (Morse et Hedlund, 1938; Brlek, 1989; Allouche et Shallit, 1999). Ses premières lettres sont :

*abbabaabbaababbabaababbaabbabaabbaababbaabbabaababbabaabbaababba...*

Un *point fixe d'un morphisme  $\psi$*  est un mot fini ou infini  $w$  tel que  $\psi(w) = w$ .

Un morphisme  $\psi$  est dit *non-effaçant* si, pour toute lettre  $\alpha$ ,  $\psi(\alpha) \neq \varepsilon$ . Un tel morphisme est également un morphisme du demi-groupe  $\Sigma^+$  vers lui-même.

**Miroir et palindromes.** On définit le *miroir* d'un mot fini  $w$  par  $\rho(w) = w_{|w|} \dots w_2 w_1$ . Un *palindrome* est un mot fini qui satisfait l'équation  $w = \rho(w)$ .

*Exemple 0.0.7.* Les mots de la langue française *kayak*, *serres* et *ressasser* sont des palindromes. La phrase de langue anglaise « A man, a plan, a canal: Panama! » est aussi un palindrome, si on fait fi des espaces, de la casse et des symboles de ponctuation.

L'ensemble de tous les palindromes d'un mot est noté  $\text{Pal}(w)$ . C'est un sous-ensemble de  $\mathcal{L}(w)$ . La cardinalité de  $\text{Pal}(w) \cap \mathcal{L}_n(w)$  est la *complexité palindromique*, ici abrégée par  $\mathcal{P}(w, n)$ .

### Antimorphismes et pseudopalindromes

Un *antimorphisme* est une fonction entre monoïdes qui préserve l'opération tout en renversant l'ordre :

$$\begin{aligned} \theta : M &\rightarrow M' \\ uv &\mapsto \theta(v)\theta(u). \end{aligned}$$

On peut notamment choisir le monoïde libre  $\Sigma^*$ . L'antimorphisme est donc la composition d'un morphisme et du miroir :

$$\begin{aligned} \theta : \Sigma^* &\rightarrow \Sigma^* \\ w_1 \dots w_n &\mapsto \theta(w_n) \dots \theta(w_1). \end{aligned}$$

*Exemple 0.0.8.* Le miroir est un antimorphisme :

$$\begin{aligned} \rho : \Sigma^* &\rightarrow \Sigma^* \\ w_1 \dots w_n &\mapsto \text{Id}(w_n) \dots \text{Id}(w_1). \end{aligned}$$

Le miroir a une propriété supplémentaire : c'est une *involution*. En effet, renverser deux fois l'ordre des lettres retourne le mot initial :  $\rho^2(w) = w$ .

Un antimorphisme non-effaçant ne peut avoir cette propriété que si l'image d'une lettre par le morphisme est toujours une lettre; on dit alors que l'antimorphisme est strictement alphabétique. Autrement, le nombre de lettres dans le mot augmente chaque fois qu'on applique le morphisme.

Finalement, pour que le morphisme soit involutif, c'est-à-dire que deux applications du morphisme sur un mot retourne le mot initial, il faut que le morphisme soit une permutation des lettres d'ordre deux.

**Proposition 0.0.9** (Folklore). *Soit  $\theta : \Sigma^* \rightarrow \Sigma^*$  un antimorphisme involutif. Alors, il existe une permutation  $\sigma$  sur l'alphabet  $\Sigma$  telle que  $\sigma^2 = \text{Id}$  et  $\theta = \rho \circ \sigma = \sigma \circ \rho$ .*

Lorsque le contexte est clair, on note  $\hat{w} = \theta(w)$  et  $\tilde{w} = \rho(w)$  plutôt que de préciser  $\sigma$ .

*Démonstration.* Comme  $\theta$  est un antimorphisme involutif,

$$w_1 \dots w_n = \theta^2(w_1 \dots w_n) = \theta(\theta(w_n) \dots \theta(w_1)) = \theta^2(w_1)\theta^2(w_2) \dots \theta^2(w_n).$$

Ainsi,

$$\theta^2 : w_1 \mapsto w_1, \dots, w_n \mapsto w_n$$

et la restriction de  $\theta$  aux lettres de  $\Sigma$  est une involution  $\sigma : \Sigma \rightarrow \Sigma$ . D'où le morphisme composé avec l'image miroir pour obtenir  $\theta$  est  $\sigma$ , une permutation involutive.  $\square$

C'est à partir de ces définitions qu'on peut introduire les  $\sigma$ -palindromes, parfois aussi appelés pseudopalindromes (de Luca et de Luca, 2006). Les  $\sigma$ -palindromes



sont les mots tels que  $w = \widehat{w}$ . Les palindromes habituels sont donc des Id-palindromes. En général, on suppose que  $\sigma$  est une involution fixée.

*Exemple 0.0.10.* Soit l'alphabet  $\{e, n, t\}$  et considérons la permutation  $\sigma : n \leftrightarrow t, e \leftrightarrow e$ . Le mot *entente* est un  $\sigma$ -palindrome.

*Exemple 0.0.11.* Considérons maintenant  $\sigma : l \leftrightarrow v$ . Le mot *élevé* est un  $\sigma$ -palindrome.

*Exemple 0.0.12.* Si on pose  $\sigma : \{b \leftrightarrow r, c \leftrightarrow d\}$ , le mot *abracadabra* est un  $\sigma$ -palindrome.

Comme pour les palindromes, l'ensemble des pseudopalindromes est noté  $\text{Pal}_\sigma(w)$  et le nombre de facteurs  $\sigma$ -palindromes de longueur  $n$  dans un mot  $w$  est appelé *complexité  $\sigma$ -palindromique* :  $\mathcal{P}_\sigma(w, n) = \#(\mathcal{L}_n(w) \cap \text{Pal}_\sigma(w))$ .

**Mots conjugués.** Deux mots  $u$  et  $v$  sont dits *conjugués* s'il existe des mots  $x$  et  $y$  tels que  $u = xy$  et  $v = yx$ . Ce résultat classique est abondamment utilisé dans ce mémoire :

**Proposition 0.0.13** (Proposition 1.3.4, (Lothaire, 1983); lemme 2.10, (Labbé, 2008)). *Soit  $x, z \in \Sigma^+$  et  $y \in \Sigma^*$  tels que  $xy = yz$ . Alors, il existe deux mots  $u$  et  $v$  et un entier  $k$  tels que*

$$x = uv, \quad z = vu, \quad y = u(vu)^k.$$

*De plus, si  $xy$  est un  $\sigma$ -palindrome, il en est de même pour  $u$  et  $v$ .*

Deux cas particuliers de cette proposition sont :

**Proposition 0.0.14** (Lemme 2.11, (Labbé, 2008)). *Soit  $y$  un  $\sigma$ -palindrome qui se chevauche lui-même :  $xy = yz$  avec  $|y| \geq |x|$ . Alors  $xy$  est un  $\sigma$ -palindrome.*

**Proposition 0.0.15** (Proposition 1.3.2, (Lothaire, 1983)). *Deux mots  $u$  et  $v$  commutent, c'est-à-dire que  $uv = vu$ , si et seulement s'ils sont puissances d'un même mot.*

Finalement, on introduit une dernière proposition sur la périodicité des mots qui se chevauchent.

**Proposition 0.0.16.** *Soit  $w = xy = x'y'$  avec  $x, y, x', y'$  des  $\sigma$ -palindromes. Si  $w$  est primitif, alors  $x = x'$  et  $y = y'$ .*

*Démonstration.* Si  $x \neq x'$ , on peut supposer sans perte de généralité que  $|x| > |x'|$ . Il est alors facile de voir qu'il existe un mot  $z$  tel que  $x = x'z$  et  $y' = zy$  (lemme de Levi). Comme  $x$  et  $x'$  sont des  $\sigma$ -palindromes,

$$x'z = x = \widehat{x} = \widehat{z\widehat{x}'} = \widehat{z}x'.$$

Par la proposition 0.0.13,  $z$  est aussi un  $\sigma$ -palindrome et  $x'z = zx$ . D'où  $x'$  et  $z$  sont puissances d'un même mot primitif  $u$  (proposition 0.0.15). De plus, comme  $y' = zy$ , on fait le même raisonnement pour trouver que  $yz = zy$  et alors,  $y$  et  $z$  sont des puissances de  $u$ . On contredit ainsi la primitivité de  $w$ . Donc  $x = x'$  et  $y = y'$ .  $\square$

**Mots de retour complet.** Soit  $w$  un mot et soit  $u$  un de ses facteurs. On définit l'ensemble des *mots de retour complet de  $u$  dans  $w$*  par l'ensemble

$$R(w, u) = \left\{ \begin{array}{l} v \in \mathcal{L}(w) \text{ débute et termine par } u \text{ sans autre occurrence de } u, \\ |v| > |u| \end{array} \right\}.$$

*Exemple 0.0.17.* Lorsqu'on se place dans le contexte du mot *mentalement*, quelques-uns de ses mots de retour complet sont dans ce tableau :

$u$	$R(\text{mentalement}, u)$
ment	{mentalement}
m	{mentalem}
e	{entale, eme}

**Tableau 0.1** Quelques mots de retour complet de facteurs du mot *mentalement*.

De manière analogue, pour toute involution  $f$  sur un mot (pas nécessairement un morphisme), on définit le *mot de retour complet de l'ensemble  $\{u, f(u)\}$  dans  $w$*  comme l'ensemble  $R(w, \{u, f(u)\})$  des mots qui débutent et terminent par  $u$  ou  $f(u)$  sans autre occurrence de  $u$  ou  $f(u)$ . Pour simplifier la notation, on note  $R(w, u)$  pour  $R(w, \{u\})$  et, si le contexte est clair, on dit que les mots de retour complet de  $\{u, \theta(u)\}$  sont les *mots de  $\theta$ -retour complet de  $u$* .

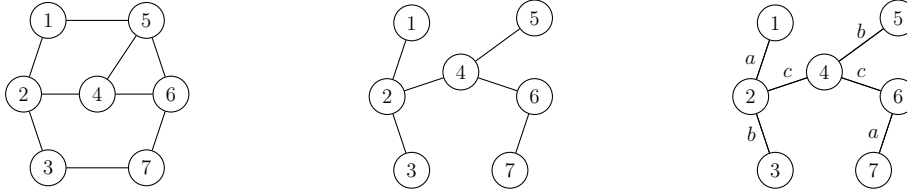
*Exemple 0.0.18.* Un mot de retour complet de  $\{u, \tilde{u}\}$  dans  $w$  est un facteur de  $w$  qui commence et termine par  $u$  ou  $\tilde{u}$  sans autre occurrence de  $u$  ou de  $\tilde{u}$ .

*Exemple 0.0.19.* Le mot *alabama* avec la permutation  $\sigma : l \leftrightarrow m$  est tel que  $R(\text{alabama}, \{la, \widehat{la}\}) = R(\text{alabama}, \{la, am\}) = \{\text{labam}\}$ .

Enfin, on remarque que la propriété de récurrence est essentielle pour que l'ensemble des mots de retour complet de chaque facteur soit non-vidé. De plus, la récurrence uniforme est nécessaire pour que l'ensemble des mots de retour complet d'un langage donné soit fini.

## Graphes et arbres

Un *graphe simple* est une structure définie par un ensemble de *sommets*  $S$  et un ensemble d'*arêtes*  $A \subseteq \{\{s_1, s_2\} \in S^2 \mid s_1 \neq s_2\}$ .



**Figure 0.1** Un graphe simple, avec (à gauche) et sans (au centre) cycle. Ses sommets sont  $\{1, 2, 3, 4, 5, 6, 7\}$  et les arêtes du graphe au centre sont  $\{\{1, 2\}, \{2, 3\}, \{2, 4\}, \{4, 5\}, \{4, 6\}, \{6, 7\}\}$ . Sur la droite, il est étiqueté par l'alphabet  $\{a, b, c\}$ .

Une *chaîne simple* dans un graphe est définie par une suite ordonnée d'arêtes adjacentes dans laquelle chaque sommet est présent au plus une fois. Dans le cas d'un graphe orienté, on parlera plutôt de *chemin simple*. Un *cycle* est une chaîne qui débute et termine au même sommet. Si les arêtes du graphe sont étiquetées par des lettres de l'alphabet (ou des mots), on définit la *trace* d'une chaîne comme la concaténation des lettres sur ses arêtes.

Un *arbre* est un graphe connexe, c'est-à-dire qu'il existe une chaîne entre chaque paire de sommets, qui ne contient aucun cycle. La proposition suivante est bien connue.

**Proposition 0.0.20.** *Un graphe simple est un arbre si et seulement si, pour tout couple de sommets  $(s_1, s_2) \subseteq S^2$ , il existe une unique chaîne du sommet  $s_1$  vers le sommet  $s_2$ .*

Les sommets d'un graphe sont appelés *nœuds* si le graphe est un arbre. Le *degré*

d'un nœud est donné par le nombre d'arêtes qui lui sont adjacentes. Une *feuille* est un nœud de degré 1.

**Langage d'un arbre** On considère un arbre  $T$  dont les arêtes sont étiquetées par l'alphabet  $\Sigma$ . Comme il existe dans un arbre une unique chaîne simple entre n'importe quel couple de nœuds, la fonction  $p(x, y)$  qui retourne la suite des arêtes le long de la chaîne du nœud  $x$  au nœud  $y$  est bien définie, tout comme la suite  $\pi(x, y)$  qui est la concaténation des étiquettes de  $p(x, y)$ , sa trace.

Par analogie avec les mots,  $\pi(x, y)$  est appelé *facteur* de l'arbre  $T$  et l'ensemble des facteurs de  $T$  est son *langage* :

$$\mathcal{L}(T) = \{\pi(x, y) \mid x, y \in \text{nœuds}(T)\}.$$

On définit le langage palindrome de  $T$  par

$$\text{Pal}(T) = \{w \in \mathcal{L}(T) \mid w = \tilde{w}\}.$$

Bien que la *taille d'un arbre*  $T$  soit généralement définie par son nombre de nœuds, on utilise ici cette expression pour désigner le nombre d'arêtes. Cette notation met l'accent sur la relation avec les mots, où la longueur est définie comme le nombre de lettres (en comptant toutes les occurrences).

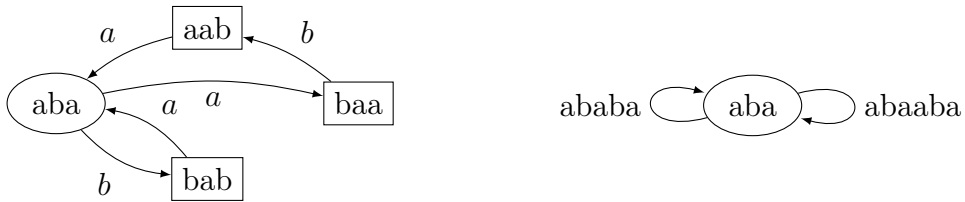
On note que, comme chaque chaîne non-vide est déterminée par les arêtes initiale et finale, la cardinalité du langage de  $T$  est bornée :

$$|\mathcal{L}(T)| \leq |T|^2 + 1. \tag{0.1}$$

**Graphes de Rauzy** Le graphe de Rauzy  $\Gamma_n$  d'un mot  $w$  est un graphe orienté. Il a pour sommets  $\mathcal{L}_n(w)$  et il existe une arête étiquetée par la lettre  $a$  du sommet

indexé par  $u$  à celui indexé par  $v$  si et seulement si  $u = u_1 \dots u_n$  et  $v = u_2 \dots u_n a$ . Le graphe de Rauzy diffère légèrement du graphe simple, puisque les boucles, c'est-à-dire les arêtes d'un sommet vers lui-même, sont admises et puisqu'il est orienté.

Les graphes de Rauzy sont utiles pour étudier la complexité en facteurs puisque  $\mathcal{C}(n)$  est égal au nombre de sommets et  $\mathcal{C}(n+1)$  au nombre d'arêtes. Une version alternative, le graphe de Rauzy réduit, a été élaborée (Baláži *et al.*, 2007) : il est construit de la même façon que le graphe de Rauzy original, mais seuls les facteurs spéciaux sont considérés. En effet, les autres facteurs correspondent à des sommets n'ayant que deux voisins et les chemins dont ils font partie sont complètement déterminés. Il diffère de plus du graphe de Rauzy original par l'étiquetage des arêtes : une arête de  $u$  à  $v$  a pour préfixe  $u$  et pour suffixe  $v$ .



**Figure 0.2** Graphe de Rauzy  $\Gamma_3(\text{Fib})$  (à gauche) et graphe de Rauzy réduit  $\Gamma'_3(\text{Fib})$  (à droite) du mot de Fibonacci.

### Notations asymptotiques

Pour décrire le comportement asymptotique des fonctions, on utilise les notations  $\Theta$ ,  $\mathcal{O}$  et  $\Omega$ .

On dit d'une fonction  $f$  qu'elle appartient à l'ensemble  $\mathcal{O}(g(n))$  s'il existe des constantes positives  $c$  et  $N$  telles que

$$f(n) \leq c \cdot g(n), \quad \forall n > N.$$

De la même façon, on dit que  $f$  est dans l'ensemble  $\Omega(g(n))$  s'il existe des constantes positives  $c$  et  $N$  telles que

$$f(n) \geq c \cdot g(n), \quad \forall n > N.$$

Enfin, on dit que  $f(n) \in \Theta(g(n))$  si  $f(n)$  est à la fois dans  $\Omega(g(n))$  et dans  $\mathcal{O}(g(n))$  :

$$\exists c_1, c_2, N \text{ tels que } c_1 \cdot g(n) \leq f(n) \leq c_2 \cdot g(n), \quad \forall n > N.$$

Ces notations désignent le comportement asymptotique d'une fonction. Leur utilisation vise souvent à simplifier les calculs, ou la lecture, en ne s'attardant qu'à l'ordre de croissance d'une fonction. On peut, notamment, avec ces notations, ne se soucier que des termes d'ordre supérieur dans une somme.

*Exemple 0.0.21.* Soit le polynôme  $P(x) = a_n x^n + a_{n-1} x^{n-1} + a_{n-2} x^{n-2} + \dots + a_1 x + a_0$ . Alors,  $P(x) \in \Theta(x^n)$ . Il est aussi vrai, mais moins précis, que  $P(x) \in \Omega(x^{n-1})$  et  $P(x) \in \mathcal{O}(e^x)$ .

*Exemple 0.0.22.* La fonction logarithmique  $\ln(4x)$  appartient aux ensembles  $\Theta(\log(x))$ ,  $\mathcal{O}(\sqrt{x})$  et  $\Omega(1)$ .





## CHAPITRE I

### COMPLEXITÉ ET DÉFAUT PALINDROMIQUES DES MOTS INFINIS.

Ce premier chapitre présente des façons de dénombrer les  $\sigma$ -palindromes pour des mots arbitrairement longs et infinis. Il introduit pour ce faire une nouvelle mesure, le *défait  $\sigma$ -palindromique*.

#### 1.1 Nombre de palindromes et défaut

Un théorème bien connu sur les palindromes porte sur leur nombre maximal en tant que facteurs dans un mot :

**Théorème 1.1.1** (Proposition 1, (Droubay *et al.*, 2001)). *Le nombre de palindromes distincts dans un mot  $w$  est au plus  $|w| + 1$ .*

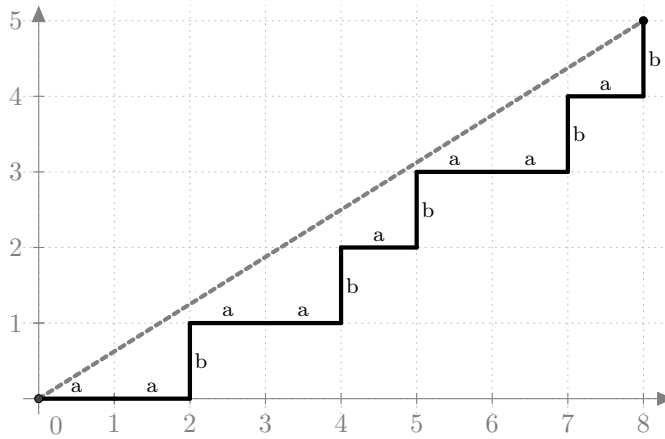
Il est aussi connu que plusieurs mots finis atteignent cette borne, comme en témoigne l'exemple ci-dessous. Ces mots font l'objet de remarques aux sections 1.2 et 1.3.

*Exemple 1.1.2.* Les *mots de Christoffel* sont des mots qui codent les segments de droites discrètes de pente rationnelle (voir figure 1.1). Ces mots ont aussi la propriété d'avoir le nombre maximal de palindromes. Par exemple, le mot de

Christoffel associé au couple d'entiers premiers entre eux  $(5, 8)$  est  $aabaababaabab$ .

Il est de longueur 13 et ses palindromes sont :

$$\left\{ \varepsilon, a, b, aa, aba, bab, baab, aabaa, ababa, abaaba, aababaa, babaabab, baababaab, abaababaaba \right\}.$$



**Figure 1.1** Le mot de Christoffel qui code la droite de pente  $\frac{5}{8}$ . Les mots de Christoffel sont construits en prenant le chemin en dessous de la droite, mais le plus près de celle-ci possible, avec uniquement des pas horizontaux ou verticaux. Pour une pente  $\frac{a}{b}$  (en fraction réduite), le mot de Christoffel associé est de longueur  $a + b$ .

On peut également étudier ce qui se passe dans certains mots infinis en s'attardant à leurs facteurs. D'une part, il existe des mots infinis dont chacun des préfixes contient le nombre maximal de palindromes. Certains mots périodiques constituent un tel exemple de mots contenant *beaucoup* de palindromes.

*Exemple 1.1.3.* Considérons le mot infini  $(ab)^\omega$ . Ses palindromes sont tous de

longueur impaire, à l'exception du mot vide, et sont les suivants :

$$\{\varepsilon, a, b, aba, bab, ababa, babab, abababa, bababab, ababababa, babababab, \dots\}$$

Non seulement il existe une infinité de tels palindromes, mais en plus, pour n'importe quel préfixe de  $(ab)^\omega$ , le plus long palindrome suffixe est soit le préfixe au complet (s'il termine par  $a$ ), soit le préfixe auquel on a retiré la première lettre (s'il termine par  $b$ ). Il y a donc autant de lettres que de palindromes non-vides.

D'autre part, il existe des mots contenant une infinité de palindromes, mais qui n'ont pas cette propriété. Le mot de Thue-Morse en est un exemple :

*Exemple 1.1.4.* On peut dénombrer les palindromes distincts dans le préfixe de longueur 11 du mot de Thue-Morse (exemple 0.0.6). Il en contient 10 :

$$\{\varepsilon, a, b, aa, bb, aba, bab, abba, baab, aabbaa\}.$$

Il n'est pas difficile de voir que  $\mu^{2^n}(a)$  est un palindrome pour tout entier  $n$ . Ainsi, le mot de Thue-Morse contient une infinité de palindromes. Il a cependant moins de palindromes que de lettres, contrairement au mot périodique  $(ab)^\omega$ , par exemple.

Afin de quantifier cette propriété, Srećko Brlek, Sylvie Hamel, Maurice Nivat et Christophe Reutenauer ont introduit la notion de *défaut palindromique* pour les mots finis. Il s'agit de la différence entre le nombre maximal de palindromes qu'un mot  $w$  peut contenir (c'est-à-dire  $|w| + 1$ , par le théorème 1.1.1) et le nombre de palindromes qu'il contient effectivement. Plus précisément, le défaut palindromique d'un mot fini  $w$  est donné par la formule, présentée dans (Brlek *et al.*,

2004),

$$\mathcal{D}(w) = |w| + 1 - |\text{Pal}(w)|.$$

## 1.2 $\sigma$ -Défaut des mots finis

De la même façon que les palindromes se généralisent aux  $\sigma$ -palindromes, on peut étendre la définition du défaut pour quantifier la saturation en  $\sigma$ -palindromes. Pour ce faire, on doit établir le nombre maximal de  $\sigma$ -palindromes qu'un mot peut contenir.

**Proposition 1.2.1** (Proposition 8, (Starosta, 2011)). *Pour tout mot fini  $w$ , soit  $k$  le nombre de transpositions de  $\sigma$  dont au moins une des lettres apparaît dans  $w$ . Alors,*

$$|\text{Pal}_\sigma(w)| \leq |w| + 1 - k.$$

On peut d'ailleurs remarquer que le théorème 1.1.1 est un corollaire de la proposition;  $k$  vaut alors 0.

*Exemple 1.2.2.* Pour voir que des mots atteignent la borne de la proposition 1.2.1, on peut par exemple étudier le mot *ombrelle* pour  $\sigma : \{o \leftrightarrow r, m \leftrightarrow b, e \leftrightarrow e, l \leftrightarrow l\}$ . Les  $\sigma$ -palindromes de *ombrelle* sont alors :

$$\{\varepsilon, e, l, ll, mb, elle, ombr\}.$$

Ce mot contient donc 7  $\sigma$ -palindromes et 8 lettres et la permutation  $\sigma$ , 2 transpositions.

Avant de procéder à la démonstration, on définit une  $\sigma$ -lacune d'un mot  $w$ . Soit  $i$  un entier strictement positif. Soit  $u$  le plus long suffixe de  $w_1 \dots w_i$  qui est un

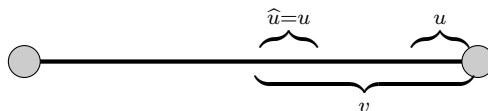
$\sigma$ -palindrome. On dit que  $i$  est une  $\sigma$ -lacune si  $u$  apparaît dans  $w_1 \dots w_{i-1}$ .

*Exemple 1.2.3.* Le mot *institution* comporte les lacunes 5 et 11 :

		1	2	3	4	5	6	7	8	9	10	11
		i	n	s	t	i	t	u	t	i	o	n
PLPS	0	1	1	1	1	*	3	1	3	5	1	*

**Tableau 1.1** Lacunes et longueurs des plus longs  $\sigma$ -palindromes suffixes des préfixes d'*institution*. Les lacunes sont identifiées par des astérisques.

*Démonstration de la proposition 1.2.1.* Pour toute lettre  $a$  et pour tout mot  $u$ ,  $|\text{Pal}_\sigma(ua)| \leq |\text{Pal}_\sigma(u)| + 1$ . Autrement, deux  $\sigma$ -palindromes terminant par la dernière occurrence de  $a$  seraient ainsi créés et le plus court est certainement préfixe du plus long.



**Figure 1.2** L'ajout d'une lettre à un mot crée au plus un  $\sigma$ -palindrome original.

De plus,  $|\text{Pal}_\sigma(\varepsilon)| = 1$ . Alors le nombre de  $\sigma$ -palindromes dans un mot  $w$  est égal au nombre de préfixes  $u$  de  $w$  pour lesquels le préfixe de  $w$  de longueur  $|u| + 1$  contient un  $\sigma$ -palindrome de plus : ce nombre vaut au plus  $|w| + 1$ .

Finalement, pour toute transposition  $a \leftrightarrow b$  dans  $\sigma$ , la première occurrence de  $a$  ou de  $b$  est une  $\sigma$ -lacune. Supposons que  $a$  apparaît pour la première fois avant  $b$ . Alors, le  $\sigma$ -palindrome terminant à la première occurrence de  $a$  doit débiter par  $b$ ,

qui n'a pas d'occurrence dans le préfixe de  $w$ . Donc pour toute transposition dont au moins une des lettres apparaît dans le mot, il existe au moins une  $\sigma$ -lacune, ce qui termine la preuve.  $\square$

*Exemple 1.2.4* (Proposition 2, (Blondin Massé, 2008)). La seule permutation qui ne soit pas l'identité sur l'alphabet  $\{a, b\}$  échange les deux lettres; notons-la  $E$ . Elle a une unique transposition ( $a \leftrightarrow b$ ) et donc, pour tout mot,  $w \in \{a, b\}^+$ ,  $|\text{Pal}_E(w)| \leq |w|$ . Notons par ailleurs que  $\text{Pal}_E(\alpha^k) = \{\varepsilon\}$  pour toute lettre  $\alpha$ , et tout  $k \in \mathbb{N}$ .

On définit le  $\sigma$ -défaut comme la différence

$$\mathcal{D}_\sigma(w) = |w| + 1 - |\text{Pal}_\sigma(w)|.$$

### 1.2.1 Un algorithme pour calculer le $\sigma$ -défaut

Comme l'explique en partie la preuve de la proposition 1.2.1, pour tout mot fini, la différence entre sa longueur plus 1 et le nombre de  $\sigma$ -palindromes qu'il contient, autrement dit, son  $\sigma$ -défaut, est le nombre de  $\sigma$ -lacunes. On peut donc imaginer un algorithme qui calcule le défaut en comptant le nombre de  $\sigma$ -lacunes, comme l'ont présenté Alexandre Blondin Massé, Srećko Brlek, Ariane Garon et Sébastien Labbé pour  $\sigma = \text{Id}$  (Blondin Massé *et al.*, 2008b). Cet algorithme procède en deux étapes : pour chaque préfixe  $u$  du mot, on trouve son plus long  $\sigma$ -palindrome suffixe, qu'on note  $\text{PL}\sigma\text{PS}(u)$ . C'est le seul nouveau  $\sigma$ -palindrome qui peut apparaître (voir, par exemple, la preuve de la proposition 1.2.1). Ensuite, on examine si ce palindrome est uni-occurent, c'est-à-dire s'il apparaît ailleurs dans le préfixe.

---

**Algorithme 1.1** Plus long  $\sigma$ -palindrome suffixe d'un mot fini  $w$

---

```

1: fonction PL $\sigma$ PS( $w$ )
2:   pour  $i$  de 1 à  $|w|$  faire
3:     si  $w_i \dots w_{|w|} = \theta(w_i \dots w_{|w|})$  alors
4:       retourner  $w_i \dots w_{|w|}$ 
5:     fin si
6:   fin pour
7:   retourner  $\varepsilon$ 
8: fin fonction

```

---



---

**Algorithme 1.2** Le  $\sigma$ -défaut d'un mot fini  $w$

---

```

1: fonction  $\mathcal{D}_\sigma(w)$ 
2:    $d \leftarrow 0$ 
3:   pour  $i$  de 1 à  $|w|$  faire
4:      $s \leftarrow \text{PL}\sigma\text{PS}(w_1 \dots w_i)$ 
5:     si  $s$  est un facteur de  $w_1 \dots w_{i-1}$  alors
6:        $d \leftarrow d + 1$ 
7:     fin si
8:   fin pour
9:   retourner  $d$ 
10: fin fonction

```

---

Cet algorithme n'est pas optimal. En effet, sa complexité est  $\mathcal{O}(|w|^3)$  si la ligne 5 de l'algorithme pour le  $\sigma$ -défaut est implémentée de façon optimale, c'est-à-dire avec un arbre préfixe ou une bonne fonction de hachage. En revanche, on peut construire un algorithme en  $\mathcal{O}(|w|^2)$ , sous la même condition, en adoptant le principe élaboré par Johan Jeuring (Jeuring, 2013).

Le dernier algorithme permet de calculer le  $\sigma$ -défaut des mots finis et de déduire

le résultat :

**Proposition 1.2.5** (Lemme 1, (Blondin Massé *et al.*, 2008b)). *Soit  $w$  un mot fini et soit  $u$  un facteur de  $w$ . Alors,  $\mathcal{D}_\sigma(u) \leq \mathcal{D}_\sigma(w)$ .*

Bien que ces algorithmes ne fonctionnent que pour les mots finis, on peut s'en servir pour calculer le  $\sigma$ -défaut des préfixes finis des mots infinis, par exemple. La prochaine section détaille la façon de faire.

### 1.3 $\sigma$ -Défaut des mots infinis

Ainsi, la propriété énoncée plus haut pour  $(ab)^\omega$  (exemple 1.1.3), c'est-à-dire que chacun de ses préfixes contient le nombre maximal de palindromes, peut se traduire par « le défaut de chaque préfixe de  $(ab)^\omega$  est nul ».

On peut étendre la notion de défaut aux mots infinis. Pour un tel mot  $\mathbf{w}$ , son défaut est la limite des défauts de ses préfixes :

$$\mathcal{D}_\sigma(\mathbf{w}) = \lim_{n \rightarrow \infty} \mathcal{D}_\sigma(\mathbf{w}_1 \mathbf{w}_2 \dots \mathbf{w}_n).$$

Les mots de défaut nul ont été largement étudiés, et diverses terminologies ont été employées : ils ont principalement été appelés pleins (Brlék *et al.*, 2004; Blondin Massé *et al.*, 2008a; Blondin Massé *et al.*, 2008b) et riches (Bucci *et al.*, 2009; Glen *et al.*, 2009; Balková *et al.*, 2011b; Starosta, 2011; Balková *et al.*, 2013). Les prochaines pages étudient le  $\sigma$ -défaut des mots infinis. Les mots à l'étude sont les mots périodiques, les mots sturmiens et le mot de Thue-Morse.



### 1.3.1 $\sigma$ -Défaut des mots périodiques

Les mots périodiques sont des mots dont les propriétés combinatoires sont plus faciles à étudier. C'est un cas particulier de points fixes de morphismes (celui où chaque lettre a pour image un préfixe du point fixe de longueur un multiple de la période) et de mots uniformément récurrents.

Les palindromes des mots périodiques ont été largement étudiés dans un article (Brlek *et al.*, 2004) où les auteurs et auteure ont donné une caractérisation simple des mots ayant un nombre infini de palindromes. Celle-ci s'étend, par le même raisonnement, aux  $\sigma$ -palindromes :

**Théorème 1.3.1** (Théorème 4, (Brlek et Lafrenière, 2014)). *Soit  $\sigma$  une involution  $\Sigma \rightarrow \Sigma$  et  $w$  un mot primitif non-vide. Les deux conditions suivantes sont équivalentes :*

1.  $w$  est le produit de deux  $\sigma$ -palindromes;
2.  $\text{Pal}_\sigma(w^\omega)$  est infini.

*Démonstration.* 1  $\implies$  2 : Supposons que  $w = uv$  avec  $u$  et  $v$  deux  $\sigma$ -palindromes tels que  $uv \neq \varepsilon$ . Alors, pour tout entier  $n$ , le préfixe  $(uv)^n u$  de  $w^\omega$  est un  $\sigma$ -palindrome.

2  $\implies$  1 : Puisque le nombre de  $\sigma$ -palindromes est infini, il en existe qui sont arbitrairement grands. Soit  $p$  un facteur  $\sigma$ -palindrome de  $w^\omega$  de longueur supérieure à  $2|w|$ . Cela implique qu'il existe  $x$  et  $y$  tels que  $p = xw^k y$  pour un certain entier  $k \geq 1$ . Comme  $p = \widehat{p} = \widehat{y}\widehat{w}^k\widehat{x}$ ,  $\widehat{w}$  est facteur de  $ww$ , et donc il existe des facteurs  $x'$  et  $y'$  de  $w$  tels que  $ww = x'\widehat{w}y'$ , avec  $|x'| + |y'| = |w|$ . Ainsi,  $w = x'y'$  et  $y'x' = \widehat{w} = \widehat{y}'\widehat{x}'$ , et donc  $x'$  et  $y'$  sont des  $\sigma$ -palindromes.  $\square$

On note que la première condition du théorème précédent, soit d'être un produit de deux  $\sigma$ -palindromes, peut s'exprimer par rapport à la conjugaison. En effet, un produit de deux  $\sigma$ -palindromes est soit conjugué à un  $\sigma$ -palindrome ou à un produit d'un  $\sigma$ -palindrome et d'une lettre. Or, les puissances infinies d'un mot et de ses conjugués partagent les mêmes facteurs. Il n'est donc pas surprenant que cette condition apparaisse.

L'équivalence des deux énoncés permet de caractériser précisément les mots périodiques ayant un ensemble infini (respectivement, fini) de  $\sigma$ -palindromes. Afin d'ajouter un peu de précision à l'énoncé, le lemme suivant fournit les complexités en facteurs et  $\sigma$ -palindromique de ces mots.

**Lemme 1.3.2.** *Soit  $w$  un mot primitif qui soit la concaténation de deux  $\sigma$ -palindromes. Soit  $\mathbf{w} = w^\omega$ . On rappelle que  $\mathcal{C}(\mathbf{w}, n)$  désigne la complexité en facteurs de  $\mathbf{w}$  et  $\mathcal{P}(\mathbf{w}, n)$ , sa complexité palindromique. Alors pour tout  $n \geq |w|$ ,*

$$\mathcal{C}(\mathbf{w}, n) = |w| \tag{1.1}$$

et

$$\mathcal{P}_\sigma(\mathbf{w}, n) + \mathcal{P}_\sigma(\mathbf{w}, n + 1) = 2. \tag{1.2}$$

Avant de présenter la preuve, on donne deux exemples de mots périodiques satisfaisant les hypothèses du lemme.

*Exemple 1.3.3.* Regardons le mot  $(aquaplane)^\omega$  pour la permutation  $\sigma : \{a \leftrightarrow q, q \leftrightarrow l, u \leftrightarrow p, n \leftrightarrow e\}$ . Le mot primitif de la longueur de la plus petite période est le produit de deux  $\sigma$ -palindromes :  $aquapla \cdot ne$ .

Pour chaque longueur supérieure à 9, il existe dans  $(aquaplane)^\omega$  neuf facteurs de cette longueur. Ils commencent par chacun des neuf conjugués du mot  $aquaplane$ .

Comme ce mot est primitif (c'est-à-dire qu'il n'est pas la répétition d'un autre mot), ces neuf mots sont distincts.

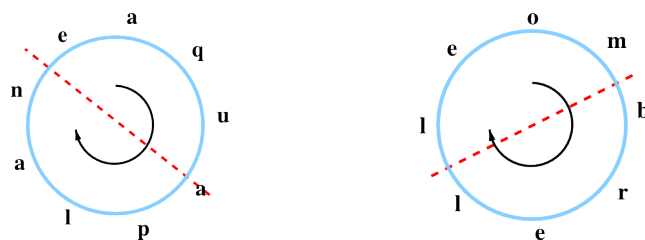
Le centre de ces  $\sigma$ -palindromes est la lettre  $a$  (ou plutôt, sa deuxième occurrence) dans le premier cas, et l'espace entre les lettres  $n$  et  $e$  dans le second. Ce sont les seuls centres de symétrie possibles. Pour le voir, on peut placer le mot sur un cercle et observer qu'il n'a qu'un axe de symétrie pour la permutation  $\sigma$  (voir figure 1.3).

Il y a donc un  $\sigma$ -palindrome de chaque longueur : pour les longueurs inférieures à 9, on peut les énumérer. Il s'agit de

$$\{\varepsilon, a, ne, uap, ane \cdot a, quapl, lane \cdot aq, aquapla, plane \cdot aqu, e \cdot aquaplan\}.$$

Pour les mots de longueur supérieure, on remarque que tous ceux qui ont comme centre la lettre  $a$  sont de longueur impaire. Il y en a un pour chaque longueur impaire (car on peut toujours ajouter une lettre de chaque côté). On fait de même pour les  $\sigma$ -palindromes qui ont pour centre l'espace entre  $n$  et  $e$  et on trouve qu'il y en a un pour chaque longueur paire. Ainsi,

$$\mathcal{P}_\sigma((aquaplane)^\omega, n) + \mathcal{P}_\sigma((aquaplane)^\omega, n + 1) = 2, \forall n \in \mathbb{N}.$$



**Figure 1.3** Les mots *aquaplane* et *ombrelle* vus comme mots circulaires. On peut observer l'unique axe de symétrie pour  $\sigma : \{q \leftrightarrow l, u \leftrightarrow p, n \leftrightarrow e\}$  et  $\sigma : \{o \leftrightarrow r, m \leftrightarrow b\}$ , respectivement.

*Exemple 1.3.4.* Considérons maintenant le mot  $(ombrelle)^\omega$  et l'involution  $\sigma : \{o \leftrightarrow r, m \leftrightarrow b, e \leftrightarrow e, l \leftrightarrow l\}$ . Le mot *ombrelle* est primitif et se décompose en deux  $\sigma$ -palindromes : *ombr · elle*.

À l'exception des lettres  $l$  et  $e$ , les  $\sigma$ -palindromes sont tous de longueur paire. On constate en effet que les seuls centres de symétrie de *ombrelle* pour  $\sigma$  sont les espaces entre les lettres  $m$  et  $b$  et entre les deux occurrences de  $l$  (voir la figure 1.3).

Pour chaque longueur paire (non-nulle), il y a donc deux  $\sigma$ -palindromes de  $(ombrelle)^\omega$  et alors,

$$\mathcal{P}_\sigma((ombrelle)^\omega, n) + \mathcal{P}_\sigma((ombrelle)^\omega, n + 1) = 2, \quad \forall n \geq 2.$$

*Démonstration du lemme 1.3.2.* L'identité (1.1) résulte de la primitivité de  $w$ . En effet, il y a au plus  $|w|$  facteurs de longueur  $n$ , étant donné que  $|w|$  est une période de  $\mathbf{w}$ . La primitivité assure que c'est la plus petite période, et qu'il y a  $|w|$  facteurs distincts de longueur  $n$ .

Il reste à observer les trois cas possibles pour les centres des palindromes (toujours une lettre, alternativement une et deux lettres, toujours deux lettres). Soit  $w = uv$  la factorisation en  $\sigma$ -palindromes telle que  $u = x\alpha\hat{x}$ ,  $v = y\beta\hat{y}$  avec  $x, y, \alpha, \beta \in \Sigma^*$ , où  $\alpha$  et  $\beta$  sont soit le mot vide ou une lettre qui est fixée par  $\sigma$ . Cette factorisation existe nécessairement (théorème 1.3.1). Ainsi, pour tout  $m \geq \frac{|w|}{2}$ ,

- si  $|\alpha| + |\beta| = 0$ , alors  $\mathcal{P}_\sigma(\mathbf{w}, 2m) = 2$  et  $\mathcal{P}_\sigma(\mathbf{w}, 2m + 1) = 0$ .
- si  $|\alpha| + |\beta| = 1$ , alors  $\mathcal{P}_\sigma(\mathbf{w}, 2m) = 1$  et  $\mathcal{P}_\sigma(\mathbf{w}, 2m + 1) = 1$ .
- si  $|\alpha| + |\beta| = 2$ , alors  $\mathcal{P}_\sigma(\mathbf{w}, 2m) = 0$  et  $\mathcal{P}_\sigma(\mathbf{w}, 2m + 1) = 2$ .

Cela permet d'observer qu'il y a au moins deux  $\sigma$ -palindromes de chaque paire de longueurs consécutives. Supposons qu'il y en ait plus : cela implique que  $w$  ait d'autres centres de symétrie que  $\alpha$  et  $\beta$  et que  $uv = xy$ , où  $x$  et  $y$  sont des  $\sigma$ -palindromes et  $x \neq u$ . Par la proposition 0.0.16,  $w$  ne peut être primitif, contredisant ainsi les hypothèses du lemme.  $\square$

Il découle du théorème 1.3.1 que les mots périodiques qui sont répétitions d'un mot qui n'est pas le produit de deux  $\sigma$ -palindromes ont un  $\sigma$ -défaut infini, puisqu'ils possèdent un nombre fini de  $\sigma$ -palindromes et une longueur infinie. On peut également calculer explicitement le  $\sigma$ -défaut des mots périodiques qui contiennent un nombre infini de  $\sigma$ -palindromes.

**Théorème 1.3.5** (Théorème 8, (Brlek et Lafrenière, 2014)). *Soit  $w = uv$  un mot primitif qui est le produit de deux  $\sigma$ -palindromes  $u$  et  $v$  tels que  $|u| \geq |v|$ , et soit  $\mathbf{w} = w^\omega$ . Alors, le  $\sigma$ -défaut de  $\mathbf{w}$  est fini et vaut :*

$$\mathcal{D}_\sigma(\mathbf{w}) = \mathcal{D}_\sigma(x), \text{ où } x \text{ est le préfixe de } \mathbf{w} \text{ de longueur } |uv| + \left\lfloor \left\lceil \frac{|u| - |v|}{3} \right\rceil \right\rfloor.$$

*Démonstration.* On doit montrer que pour n'importe quel préfixe  $p$  de  $\mathbf{w}$  de longueur supérieure à  $|uv| + \left\lfloor \left\lceil \frac{|u| - |v|}{3} \right\rceil \right\rfloor$ , le plus long  $\sigma$ -palindrome suffixe de  $p$  n'apparaît qu'une fois dans  $p$ . On rappelle que,  $u$  et  $v$  étant des  $\sigma$ -palindromes, ils sont de la forme

$$u = x\alpha\hat{x}, \quad v = y\beta\hat{y}$$

avec  $x, y, \alpha, \beta \in \Sigma^*$ ,  $\alpha$  et  $\beta$  étant soit le mot vide ou une lettre fixée par  $\sigma$ . On divise donc la preuve en trois cas selon la longueur de  $p$  :

- Si  $|p| > |uvuv|$ . Comme  $|p| > 2|w|$ , il existe un entier strictement positif  $m$

tel que :

$$p = \begin{cases} u(vu)^m vz, & \text{avec } z \text{ préfixe de } u, \\ uv(uv)^m uz, & \text{avec } z \text{ préfixe de } v. \end{cases} \quad (1.3)$$

$$(1.4)$$

Le plus long  $\sigma$ -palindrome suffixe de  $p$  est, respectivement :

$$\text{PL}\sigma\text{PS}(p) = \begin{cases} \widehat{z}(vu)^m vz, & \text{si (1.3)} \\ \widehat{z}(uv)^m uz, & \text{si (1.4)}. \end{cases}$$

Ainsi, pour tout  $|p| > 2|w|$ ,  $p = s \cdot \text{PL}\sigma\text{PS}(p)$  et  $|s| < |\text{PL}\sigma\text{PS}(p)|$ . Le plus long  $\sigma$ -palindrome suffixe est uni-occurent, car autrement, deux occurrences de  $\text{PL}\sigma\text{PS}(p)$  se chevauchent, ce qui, par la proposition 0.0.14, contredit le choix de  $\text{PL}\sigma\text{PS}(p)$ .

- Si  $|uvu| < |p| \leq |uvw|$ . Dans ce cas,  $p$  a une des formes suivantes :

$$p = \begin{cases} x\alpha\widehat{x}y\beta\widehat{y}x\alpha\widehat{x}z, & \text{avec } z \text{ préfixe de } y, \\ x\alpha\widehat{x}y\beta\widehat{y}x\alpha\widehat{x}y\beta, & \\ x\alpha\widehat{x}y\beta\widehat{y}x\alpha\widehat{x}y\beta z, & \text{avec } z \text{ préfixe de } \widehat{y}. \end{cases} \quad (1.5)$$

$$(1.6)$$

$$(1.7)$$

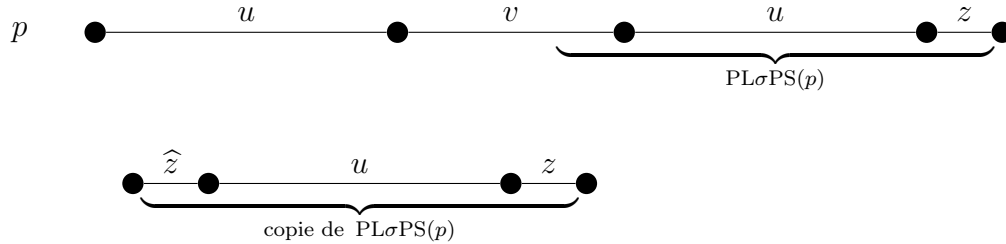
Le plus long  $\sigma$ -palindrome suffixe de  $p$  est, respectivement,

$$\text{PL}\sigma\text{PS}(p) = \begin{cases} \widehat{z}x\alpha\widehat{x}z, & \text{si (1.5)} \\ \beta\widehat{y}x\alpha\widehat{x}y\beta, & \text{si (1.6)} \\ \widehat{z}\beta\widehat{y}x\alpha\widehat{x}y\beta z & \text{si (1.7)}. \end{cases}$$

On écrit alors  $p = s \cdot \text{PL}\sigma\text{PS}(p)$ . Dans les cas (1.6) et (1.7),  $|\text{PL}\sigma\text{PS}(p)| > |s|$  et il découle de la première partie de la preuve que  $\text{PL}\sigma\text{PS}(p)$  n'apparaît qu'une seule fois.

Supposons maintenant que  $\text{PL}\sigma\text{PS}(p)$  apparaît au moins deux fois dans  $p$ . Comme  $|u| \geq |v|$ ,  $u$  doit se chevaucher lui-même. Les propositions 0.0.13 et 0.0.14, de

concert avec le fait que  $u$  est un  $\sigma$ -palindrome, impliquent alors qu'il existe deux  $\sigma$ -palindromes  $q$  et  $r$  et un entier  $k$  tels que  $u = (qr)^k q$ .



**Figure 1.4** Le chevauchement de  $\sigma$ -palindromes implique une structure particulière dans le cas représenté par l'équation (1.5).

De plus,  $rqz$  est un préfixe de  $v$  et  $\widehat{r}\widehat{q}z = \widehat{z}qr$  en est un suffixe. Ainsi,  $\widehat{z}qr \underbrace{(qr)^k q}_{u=x\alpha\bar{x}} z$  est un  $\sigma$ -palindrome suffixe de  $p$ , ce qui contredit le choix de  $\text{PL}\sigma\text{PS}(p)$ .

- Si  $|uv| + \left\lfloor \left\lceil \frac{|u|-|v|}{3} \right\rceil \right\rfloor < |p| \leq |uvu|$ . Dans ce cas,  $p = uvs$  et  $s$  est un préfixe de  $u$  de longueur supérieure à  $\left\lfloor \left\lceil \frac{|u|-|v|}{3} \right\rceil \right\rfloor$ . Soit  $r = \text{PL}\sigma\text{PS}(p)$ . Alors,

$$|r| \geq |s| + |v| + |\widehat{s}| = 2|s| + |v|.$$

Supposons qu'il existe une autre occurrence de  $r$ . Par la proposition 0.0.14 et puisque  $r$  est le plus long  $\sigma$ -palindrome suffixe, les deux occurrences de  $r$  ne se chevauchent pas :

$$|r| \leq |uvs| - 2|s| - |v| = |u| - |s|.$$

Ce qui donne le résultat suivant :

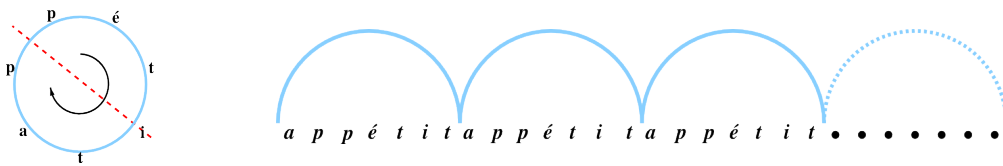
$$\begin{aligned}
 |u| - |s| &\geq 2|s| + |v| \\
 \Rightarrow 3|s| &\leq | |u| - |v| | \\
 \Rightarrow |s| &\leq \left\lfloor \left\lceil \frac{|u| - |v|}{3} \right\rceil \right\rfloor,
 \end{aligned}$$

contredisant de fait les hypothèses sur la longueur de  $s$ . □

*Exemple 1.3.6.* Posons  $\sigma : \{a \leftrightarrow \acute{e}, i, p, t\}$ , une permutation qui fixe les lettres  $i$ ,  $p$  et  $t$ . Le mot *appétit* est un produit de deux  $\sigma$ -palindromes : *appé·tit*. Si l'on observe le mot  $(\textit{appétit})^\omega$ , on se rend compte qu'il n'a qu'une  $\sigma$ -lacune, soit 1. Elle est indiquée par le symbole  $*$ .

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	...
		a	p	p	é	t	i	t	a	p	p	é	t	i	t	...
PLPS	0	*	1	2	4	1	1	3	5	7	9	11	6	8	10	...

**Tableau 1.2** Lacunes et longueurs des plus longs  $\sigma$ -palindromes suffixes des préfixes d'*appétit*.



**Figure 1.5** Le mot *appétit* et sa représentation comme mot circulaire. On peut observer l'unique axe de symétrie pour  $\sigma : \{a \leftrightarrow \acute{e}\}$ .

*Exemple 1.3.7.* En revanche, on ne peut pas toujours trouver une permutation  $\sigma$  qui fasse qu'un mot donné soit un produit de deux  $\sigma$ -palindromes. Par exemple, le mot *événement*, si l'on ignore les accents, n'admet aucune telle permutation. Autrement, le nombre plus élevé d'occurrences de  $n$  et de  $e$  impliquerait que ces deux lettres seraient fixées. Les trois involutions restantes sont :



- $\sigma : t \leftrightarrow v$ , avec  $e, m$  et  $n$  fixées. Le mot  $(\text{evenement})^\omega$  ne contient que huit  $\sigma$ -palindromes :  $\{\varepsilon, e, m, n, eme, ene, tev, nemen\}$ .
- $\sigma : m \leftrightarrow v$ , avec  $e, n$  et  $t$  fixées. Le mot périodique infini ne contient que sept  $\sigma$ -palindromes :  $\{\varepsilon, e, n, t, ene, venem, eveneme\}$ .
- $\sigma : m \leftrightarrow t$ , avec  $e, n$  et  $v$  fixées. Le mot ne contient aucun  $\sigma$ -palindrome de longueur supérieure à 3 :  $\{\varepsilon, e, n, v, ene, eve\}$ .

On peut aussi se servir du calcul des  $\sigma$ -lacunes pour vérifier si un mot est un produit de deux  $\sigma$ -palindromes. Le théorème 1.3.5 implique que si, dans  $w^\omega$ , il y a des  $\sigma$ -lacunes plus grandes que  $\frac{4}{3}|w|$ ,  $w$  n'est pas un produit de deux  $\sigma$ -palindromes. Par exemple, si on prend la permutation  $\sigma$  qui fixe  $e, n$  et  $v$  et qui échange  $m$  et  $t$  (le dernier exemple de la liste), les  $\sigma$ -lacunes d' $(\text{evenement})^\omega$  sont toutes les positions plus grandes ou égales à 6.

Enfin, la borne donnée par le théorème 1.3.5 n'est peut-être pas optimale, mais elle n'est pas loin de l'être, comme le montre l'exemple suivant :

*Exemple 1.3.8* ((Brlek et al., 2004)). Soit  $w = a^{k+1}ba^kca^kba^{k+1} \cdot c$  et  $\sigma = \text{Id}$ . Alors,  $\left\lfloor \left\lceil \frac{|u|-|v|}{3} \right\rceil \right\rfloor = \left\lfloor \frac{4}{3}(k+1) \right\rfloor$ , et  $\mathcal{D}_\sigma(w) = 1$  et  $\mathcal{D}_\sigma(w^\omega) = \mathcal{D}_\sigma(w \cdot a^k) = k+1$ .

### 1.3.2 $\sigma$ -Défaut des mots a périodiques

**Mots sturmiens.** Les *mots sturmiens* sont définis comme les mots dont la fonction de complexité vaut  $\mathcal{C}(s, n) = n + 1$  pour tous les nombres naturels. En particulier, ils sont toujours sur un alphabet binaire. D'autres définitions sont équivalentes :

- ils possèdent exactement deux facteurs palindromes de longueur impaire, et un palindrome de longueur paire, pour toute longueur (Droubay et Pirillo,

1999);

- ils codent les droites discrètes de pente irrationnelle (Morse et Hedlund, 1940). Cette définition montre qu'ils sont équivalents aux mots de billard en dimension 2 : un mot de billard de dimension  $n$  décrit la trajectoire d'une boule de billard qui frappe le côté d'un cube de dimension  $n$  avec un angle irrationnel;
- ils sont les mots équilibrés — c'est-à-dire que pour toute paire de facteurs  $\{u, v\} \subset \mathcal{L}(\mathbf{s})$ ,  $\|u|_a - v|_a\| \leq 1$  et  $\|u|_b - v|_b\| \leq 1$  — qui ne sont pas ultimement périodiques (Morse et Hedlund, 1940).

Pour plus de caractérisations, voir par exemple (Lothaire, 2002), chapitre 2.

*Exemple 1.3.9.* Le mot sturmien le plus connu est certainement le mot de Fibonacci (suite A003849 de l'OEIS). C'est aussi le point fixe du morphisme primitif  $\varphi : a \mapsto ab, b \mapsto a$ .

Fib = *abaababaabaababaababaababaababaababaababaababaababa...*

Les  $E$ -palindromes de Fib sont  $\{\varepsilon, ab, ba, abab, baba\}$ . Comme cet ensemble est fini, il est facile de voir que  $\mathcal{D}_E(\text{Fib})$  est infini.

Xavier Droubay, Jacques Justin et Giuseppe Pirillo ont démontré que chaque préfixe d'un mot sturmien contient le nombre maximal de palindromes (Droubay *et al.*, 2001), ce qui en fait une large famille d'exemples de mots infinis saturés en palindromes. Ce résultat ne peut cependant pas s'étendre aux  $\sigma$ -palindromes (théorème 6, (Blondin Massé, 2008)). Considérons la seule involution non-triviale sur un alphabet binaire :  $E : \{a \leftrightarrow b\}$ . Comme les mots sturmiens ont exactement trois facteurs de longueur 2, ils ne peuvent contenir à la fois  $aa = E(bb)$  et  $bb = E(aa)$ . Les mots sturmiens sont de plus uniformément récurrents (Morse

et Hedlund, 1940), ce qui fait que tous ses facteurs d'une certaine longueur contiennent soit  $bb$  ou  $aa$ . Par conséquent, la longueur des facteurs  $E$ -palindromes est bornée et leur nombre est fini. Le  $E$ -défaut de ces mots est donc infini, alors que le Id-défaut est nul.

**Mot de Thue-Morse.** Comme énoncé une première fois à l'exemple 1.1.4, le mot de Thue-Morse est un mot contenant un nombre infini de palindromes. Il contient aussi un nombre infini de lacunes. Il en est aussi de même pour les  $E$ -palindromes. En effet, comme mentionné plus tôt,  $\mu^{2n}(a)$  est un palindrome pour tout entier  $n$ . Alors,

$$\mu^{2n+1}(a) = \mu^{2n}(ab) = \mu^{2n}(a)\mu^{2n}(\bar{a}) = \mu^{2n}(a)\widehat{\mu^{2n}(a)}$$

et  $\mu^{2n+1}(a)$  est un  $E$ -palindrome pour tout entier  $n$ .

Il est donc facile de voir que  $\mathbf{t}$  contient à la fois une infinité de palindromes et de  $E$ -palindromes. Cependant, autant le Id-défaut que le  $E$ -défaut sont infinis. Ce mot comporte effectivement une infinité de lacunes et de  $E$ -lacunes. Alexandre Blondin Massé, Srećko Brlek et Sébastien Labbé ont d'ailleurs calculé ces positions auxquelles le plus long  $\sigma$ -palindrome suffixe n'est pas uni-occurent (Blondin Massé *et al.*, 2008c) :

**Théorème 1.3.10** (Théorème 2, (Blondin Massé *et al.*, 2008c)). *Les intervalles de  $\sigma$ -lacunes du mot de Thue-Morse sont, pour  $\sigma = \text{Id}$  :*

$$\bigcup_{n \geq 1} \left( [2^{2n+1} + 1, 5 \cdot 2^{2n-1}] \cup [3 \cdot 2^{2n+1} + 1, 13 \cdot 2^{2n-1}] \right)$$

*et pour  $\sigma = E$  :*

$$\{1, 3\} \cup \bigcup_{n \geq 1} \left( [2^{2n} + 1, 5 \cdot 2^{2n-2}] \cup [3 \cdot 2^{2n} + 1, 13 \cdot 2^{2n-2}] \right).$$

*En particulier, les intervalles de Id-lacunes et de  $E$ -lacunes sont disjoints.*

Le fait que le mot de Thue-Morse contienne une infinité de  $\sigma$ -palindromes, mais un  $\sigma$ -défaut infini, pour les deux involutions sur l'alphabet binaire a amené Edita Pelantová et Štěpán Starosta à définir les mots  $G$ -riches (Pelantová et Starosta, 2013), où  $G$  est un ensemble d'antimorphismes. Un mot  $G$ -riche contient alors le nombre maximal de  $\sigma$ -palindromes, en additionnant les valeurs pour tout  $\sigma \circ \rho \in G$ . Le mot de Thue-Morse est alors un mot  $\{\rho, E \circ \rho\}$ -riche. Cet article montre aussi que les mots uniformément récurrents qui contiennent un nombre infini de  $\sigma$ -palindromes pour chaque  $\sigma \circ \rho \in G$  (comme le mot de Thue-Morse) sont *presque  $G$ -riches*, c'est-à-dire que la différence entre le nombre maximal de pseudopalindromes (toutes permutations de  $G$  confondues) et le nombre de pseudopalindromes atteint est finie.

La définition formelle de la  $G$ -richesse est donnée à partir d'une version modifiée du graphe de Rauzy, présenté dans les préliminaires, et est omise ici étant donné qu'elle est plutôt technique.

## CHAPITRE II

### UNE IDENTITÉ DE BRLEK ET REUTENAUER.

Le second chapitre aborde les liens entre palindromes et facteurs. Plus précisément, on y présente une équation qui unit les nombres de palindromes et de facteurs et le défaut.

#### 2.1 Complexité palindromique et complexité en facteurs

À la fois la complexité palindromique et la complexité en facteurs ont été largement étudiées. Comme préambule à l'identité de Brlek et Reutenauer, nous présentons certains résultats sur celles-là et énonçons deux théorèmes les liant.

**La complexité en facteurs de certaines familles de mots** La complexité en facteurs est un concept qui a été largement étudiée, notamment pour caractériser certaines familles de mots. Voici quelques exemples :

*Exemple 2.1.1.* Les mots sturmiens (présentés à la sous-section 1.3.2) sont justement définis comme les mots  $\mathbf{s}$  dont la fonction de complexité vaut  $\mathcal{C}(\mathbf{s}, n) = n + 1$  pour tous les nombres naturels.

*Exemple 2.1.2.* Les mots périodiques et ultimement périodiques sont les seuls mots infinis qui ont une complexité éventuellement constante (théorème 7.3, (Morse et

Hedlund, 1938)).

En effet, ils s'écrivent comme  $\mathbf{w} = u \cdot v^\omega$ ,  $u, v \in \Sigma^*$ . Leur fonction de complexité est bornée par  $|u| + |v|$ , et elle est croissante (proposition 0.0.2).

Il est de plus connu (théorème 1.3.13, (Lothaire, 2002)) que les mots aperiodiques (c'est-à-dire qu'ils ne sont pas ultimement périodiques) de plus petite complexité sont les mots sturmiens. Autrement dit, certaines fonctions de complexité ne peuvent pas être atteintes par les mots infinis, comme  $\log(n)$  ou  $\sqrt{n}$ .

*Exemple 2.1.3* ((Pansiot, 1984)). Soit  $\mathbf{w}$  un point fixe infini de morphisme. Alors il existe des constantes  $c_1$  et  $c_2$  telles que

$$c_1 f(n) \leq \mathcal{C}(\mathbf{w}, n) \leq c_2 f(n),$$

où  $f(n)$  est une des fonctions suivantes :  $1$ ,  $n$ ,  $n \log(\log(n))$ ,  $n \log(n)$  ou  $n^2$ .

*Exemple 2.1.4.* La constante de Champernowne (suite A033307 de l'OEIS) est construite par la concaténation des nombres naturels en base 10 :

$$C_{10} = 12345678910111213141516171819202122232425262728293031 \dots$$

Sa fonction de complexité vaut donc  $\mathcal{C}(C_{10}, n) = 10^n$ , car tous les nombres naturels en base 10 apparaissent dans la suite.

### Complexité palindromique

Il n'est pas surprenant que le nombre de palindromes de chaque longueur dans un mot donné ait été étudié.

À cet égard, la complexité palindromique des mots sturmiens est une de leurs caractérisations :

**Théorème 2.1.5** (Théorème 5, (Droubay et Pirillo, 1999)). *Soit  $\mathbf{w}$  un mot infini. Alors,  $\mathbf{w}$  est un mot sturmien si et seulement si, pour tout entier positif  $n$ ,*

$$\mathcal{P}(\mathbf{w}, n) = \begin{cases} 1 & \text{si } n \text{ est pair;} \\ 2 & \text{si } n \text{ est impair.} \end{cases}$$

Bien entendu, la complexité en palindromes est bornée par la complexité en facteurs, les palindromes représentant un sous-ensemble des facteurs.

Des bornes plus intéressantes ont aussi été trouvées pour certaines familles de mots. Les deux énoncés qui suivent démontrent un certain intérêt pour des liens entre complexités en facteurs et palindromique :

**Théorème 2.1.6** (Théorème 12, (Allouche *et al.*, 2003)). *Soit  $\mathbf{w}$  un mot infini qui n'est pas ultimement périodique. Alors, pour tout  $k \geq 1$ ,*

$$\mathcal{P}(\mathbf{w}, k) < \frac{16}{k} \mathcal{C} \left( \mathbf{w}, k + \left\lfloor \frac{k}{4} \right\rfloor \right).$$

**Lemme 2.1.7** (Théorème 1, (Starosta, 2011)). *Soit  $\mathbf{w}$  un mot dont le langage est fermé par  $\theta$ . Alors,*

$$\mathcal{P}_\sigma(\mathbf{w}, n) + \mathcal{P}_\sigma(\mathbf{w}, n + 1) \leq \Delta \mathcal{C}(\mathbf{w}, n) + 2, \quad \forall n \in \mathbb{N},$$

où  $\Delta \mathcal{C}(\mathbf{w}, n) = \mathcal{C}(\mathbf{w}, n + 1) - \mathcal{C}(\mathbf{w}, n)$ .

Le dernier lemme avait été démontré quelques années avant dans le cas où  $\sigma = \text{Id}$  (Baláži *et al.*, 2007).

## 2.2 L'identité en question

Suivant les définitions déjà vues, on considère maintenant la relation entre les complexités en facteurs et  $\sigma$ -palindromique d'un mot. Plus précisément, on

étudie une relation proposée par Srečko Brlek et Christophe Reutenauer (Brlek et Reutenauer, 2011) qui utilise le défaut pour quantifier la différence entre la complexité palindromique et la complexité en facteurs.

L'équation en question est donnée par

$$2\mathcal{D}_\sigma(w) = \sum_{n=0}^{\infty} \Delta\mathcal{C}(w, n) + 2 - \mathcal{P}_\sigma(w, n) - \mathcal{P}_\sigma(w, n+1), \quad (2.1)$$

où  $\Delta\mathcal{C}(w, n) = \mathcal{C}(w, n+1) - \mathcal{C}(w, n)$ .

Elle a été démontrée pour  $\sigma = \text{Id}$  dans le cas des mots finis et périodiques par Srečko Brlek et Christophe Reutenauer (Brlek et Reutenauer, 2011), alors que L'ubomíra Balková, Edita Pelantová et Štěpán Starosta ont montré que les mots uniformément récurrents (Balková *et al.*, 2011b; Balková *et al.*, 2011a) et ceux dont le langage est fermé par miroir (Balková *et al.*, 2013) satisfont aussi l'équation (2.1) quand  $\sigma = \text{Id}$ .

Une question naturelle semble être d'étendre le théorème aux involutions  $\sigma \neq \text{Id}$ . C'est ce que fait ce chapitre, dans tous les cas déjà connus.

### 2.3 Mots finis

**Théorème 2.3.1.** *Pour tout mot  $w$  fini et pour toute involution  $\sigma$  sur l'alphabet  $\Sigma$ ,*

$$2\mathcal{D}_\sigma(w) = \sum_{n=0}^{|w|} \Delta\mathcal{C}(w, n) + 2 - \mathcal{P}_\sigma(w, n) - \mathcal{P}_\sigma(w, n+1).$$

*Démonstration.* Le mot  $w$  étant fini, il est clair que

$$\mathcal{P}_\sigma(w, |w| + 1) = \mathcal{C}(w, |w| + 1) = 0.$$



On peut alors réécrire l'équation et utiliser les propriétés des sommes télescopiques :

$$\begin{aligned}
& \sum_{n=0}^{|w|} \Delta \mathcal{C}(w, n) + 2 - \mathcal{P}_\sigma(w, n) - \mathcal{P}_\sigma(w, n+1) \\
&= \underbrace{\mathcal{C}(w, |w|+1)}_{=0} - \underbrace{\mathcal{C}(w, 0)}_{=1} + 2(|w|+1) - \sum_{n=0}^{|w|} (\mathcal{P}_\sigma(w, n+1) \\
&\quad + \mathcal{P}_\sigma(w, n)) \\
&= -1 + 2(|w|+1) - 2 \sum_{n=0}^{|w|} \mathcal{P}_\sigma(w, n) - \underbrace{\mathcal{P}_\sigma(w, |w|+1)}_{=0} + \underbrace{\mathcal{P}_\sigma(w, 0)}_{=1} \\
&= -1 + 2(|w|+1) - 2|\text{Pal}_\sigma(w)| + 1 \\
&= 2\mathcal{D}_\sigma(w). \quad \square
\end{aligned}$$

Pour la suite de ce chapitre, on notera la partie à l'intérieur de la somme dans l'équation 2.1 comme suit :

$$T_\sigma(w, n) = \Delta \mathcal{C}(w, n) + 2 - \mathcal{P}_\sigma(w, n) - \mathcal{P}_\sigma(w, n+1).$$

*Exemple 2.3.2.* Considérons le mot  $w = \textit{institution}$ . On peut étudier la longueur du plus long palindrome suffixe de chacun de ses préfixes, comme le montre le tableau 1.2.3.

On sait donc que le défaut d'*institution* est 2. On sait aussi combien de palindromes de chaque longueur il contient. Après avoir calculé le nombre de facteurs distincts, on peut calculer les statistiques présentées au tableau 2.3.2.

Sur la base de ces calculs, on retrouve bien le résultat du théorème 2.3.1 :

$$2\mathcal{D}(\textit{institution}) = \sum_{n=0}^{|\textit{institution}|} T(\textit{institution}, n).$$

$n$	0	1	2	3	4	5	6	7	8	9	10	11
$\mathcal{C}(w, n)$	1	6	9	9	8	7	6	5	4	3	2	1
$\mathcal{P}(w, n)$	1	6	0	2	0	1	0	0	0	0	0	0
$T(w, n)$	0	-1	0	-1	0	0	1	1	1	1	1	1

**Tableau 2.1** Complexités palindromique et en facteurs du mot *institution*.

Cette propriété ne peut être étendue à tous les mots infinis (Brlek et Reutenauer, 2011). Bien que certaines propriétés, comme le  $\sigma$ -défaut, se généralisent plutôt bien en prenant la limite des facteurs, ou préfixes, des mots, un problème se pose pour la positivité des valeurs  $T_\sigma(\mathbf{w}, n)$ . Par exemple, si  $\mathbf{w}$  est un mot infini non récurrent, il est possible que pour certaines valeurs de  $n$ ,  $T_\sigma(\mathbf{w}, n)$  soit strictement négatif. Dans le cas des mots récurrents, il n'est pas assuré que ces valeurs soient positives si  $\mathbf{w}$  n'est pas uniformément récurrent ou si l'ensemble des facteurs de  $\mathbf{w}$  n'est pas fermé par  $\theta$ .

*Exemple 2.3.3.* La concaténation de la lettre  $c$  et du mot de Fibonacci sur l'alphabet  $\{a, b\}$  est un mot non-récurrent, car  $c$  n'apparaît qu'une fois.

Ses facteurs de longueur 1 sont  $a$ ,  $b$  et  $c$  et ses facteurs de longueur 2 sont  $aa$ ,  $ab$ ,  $ba$  et  $ca$ . Il contient donc un seul palindrome de longueur 2 et trois de longueur 1.

Ainsi,

$$T_{\text{Id}}(c \cdot \text{Fib}, 1) = \underbrace{\Delta\mathcal{C}(c \cdot \text{Fib}, 1)}_{=1} + 2 - \underbrace{\mathcal{P}(c \cdot \text{Fib}, 1)}_{=3} - \underbrace{\mathcal{P}(c \cdot \text{Fib}, 2)}_{=1} = -1.$$

## 2.4 Mots périodiques

Même s'il n'existe pas de preuve générale, il est possible de démontrer l'identité pour certains types de mots, à commencer par les mots périodiques. Afin de

démontrer l'identité, il est cependant nécessaire de connaître le nombre de facteurs  $\sigma$ -palindromes dans  $w^2$ .

**Lemme 2.4.1** (Analogue au lemme 4 (ii), (Brlek et Reutenauer, 2011)). *Soit  $w$  un mot primitif de la forme  $x\alpha\hat{x}\beta$ , où  $\alpha$  et  $\beta$  sont soit le mot vide, soit des lettres fixées par  $\sigma$ , et  $x$  est un mot. Le nombre de  $\sigma$ -palindromes de longueur supérieure à  $|w|$  dans  $w^2$  est alors :*

$$\sum_{n=|w|}^{2|w|} \mathcal{P}_\sigma(w^2, n) = \frac{|w| + \mathcal{P}_\sigma(w^2, |w|) + 2}{2}.$$

*Démonstration.* Une fois de plus, la preuve s'obtient de l'étude des cas possibles :

—  $|\alpha| = |\beta| = 1$ ,  $w^2 = x\alpha\hat{x}\beta x\alpha\hat{x}\beta$ ,  $\alpha, \beta \in \Sigma$ . La longueur de  $w$  est alors paire.

La lectrice ou le lecteur remarquera que, pour  $|w| + 1 < n \leq 2|w|$ ,

$$\mathcal{P}_\sigma(w^2, n) = \begin{cases} 0 & \text{si } n \text{ est pair;} \\ 1 & \text{si } n \text{ est impair.} \end{cases}$$

Pour  $n = |w| + 1$ , à la fois  $\beta x\alpha\hat{x}\beta$  et  $\alpha\hat{x}\beta x\alpha$  sont des  $\sigma$ -palindromes. Par le lemme 1.3.2, il n'y a pas de  $\sigma$ -palindrome de longueur  $|w|$ .

Ainsi,  $2 \sum_{n=|w|}^{2|w|} \mathcal{P}_\sigma(w^2, n) = |w| + 2$ .

—  $|\alpha| = 0$ ,  $|\beta| = 1$ ,  $w^2 = x\hat{x}\beta x\hat{x}\beta$ ,  $\beta \in \Sigma$ . La longueur de  $w$  est impaire.

Comme dans le premier cas, pour  $|w| \leq n \leq 2|w|$ ,

$$\mathcal{P}_\sigma(w^2, n) = \begin{cases} 0 & \text{si } n \text{ est pair;} \\ 1 & \text{si } n \text{ est impair,} \end{cases}$$

à l'exception de  $n = |w| + 1$ , qui est pair et pour lequel  $\beta x\hat{x}\beta$  est un  $\sigma$ -palindrome.

Par conséquent,  $2 \sum_{n=|w|}^{2|w|} \mathcal{P}_\sigma(w^2, n) = |w| + 3$  et il n'y a qu'un  $\sigma$ -palindrome de longueur  $|w|$ .

Ce cas est analogue à celui où  $\alpha$  est une lettre et  $\beta$  est le mot vide. Alors,  $w^2 = x\alpha\hat{x}x\alpha\hat{x}$ ,  $\alpha \in \Sigma$ .

—  $|\alpha| = 0$ ,  $|\beta| = 0$  ( $w^2 = x\hat{x}x\hat{x}$ ). La longueur de  $w$  est alors paire.

Pour  $n = |w|$ , il y a deux facteurs  $\sigma$ -palindromes, soit  $x\hat{x}$  et  $\hat{x}x$ . Pour  $|w| < n \leq 2|w|$ , il y a un facteur  $\sigma$ -palindrome pour chaque longueur paire, et aucun  $\sigma$ -palindrome de longueur impaire.

Il en découle donc que,  $2 \sum_{n=|w|}^{2|w|} \mathcal{P}_\sigma(w^2, n) = |w| + 4$  et il y a deux  $\sigma$ -palindromes de longueur  $|w|$ .  $\square$

*Exemple 2.4.2.* On peut par exemple considérer le mot *minima*, avec  $\sigma = \text{Id}$ . Le mot a bien la forme désirée (*minim*·*a*).

Les palindromes de longueur supérieure à 6 dans  $(\textit{minima})^2$  sont :

$$\{a \cdot \textit{minima}, \textit{nima} \cdot \textit{min}, \textit{inima} \cdot \textit{mini}, \textit{minima} \cdot \textit{minim}\}.$$

Il n'y a parmi eux aucun palindrome de longueur 6.

Ainsi,

$$\sum_{n=6}^{2 \cdot 6} \mathcal{P}((\textit{minima})^2, n) = \frac{|\textit{minima}| + \mathcal{P}((\textit{minima})^2, 6) + 2}{2}.$$

L'ensemble des  $\sigma$ -palindromes de longueur  $|w|$  dans  $w^2$  n'est pas toujours vide.

On aurait pu par exemple choisir le mot *minimal* et l'involution  $\sigma$  qui fixe toutes

les lettres sauf  $a$  et  $l$ , qu'elle échange. Les  $\sigma$ -palindromes de longueur au moins 7 dans  $(minimal)^2$  sont

$$\{l \cdot minima, nimal \cdot min, al \cdot minimal, inimal \cdot mini, minimal \cdot minim\}.$$

Il y a un  $\sigma$ -palindrome de longueur 7 et on trouve

$$\sum_{n=7}^{2 \cdot 7} \mathcal{P}((minimal)^2, n) = \frac{|minimal| + \mathcal{P}((minimal)^2, 7) + 2}{2} = 5.$$

Les lemmes 1.3.2 et 2.4.1 permettent de conclure avec le théorème suivant :

**Théorème 2.4.3.** *Les mots périodiques satisfont l'équation (2.1).*

*Démonstration.* Les mots finis satisfont l'équation (2.1), par le théorème 2.3.1. Pour les mots infinis, on suppose d'abord que  $\mathbf{w} = w^\omega$  contient une infinité de  $\sigma$ -palindromes. Alors,  $w$  est le produit de deux  $\sigma$ -palindromes (théorème 1.3.1) et  $\mathcal{D}_\sigma(\mathbf{w})$  est borné par  $\mathcal{D}_\sigma(w^2)$  (théorème 1.3.5). Comme pour chaque conjugué  $w'$  de  $w$ ,  $(w')^\omega$  et  $\mathbf{w}$  partagent les mêmes facteurs, on peut supposer que  $w$  est de la forme  $\alpha x \beta \hat{x}$ , où  $\alpha$  et  $\beta$  sont de longueur 0 ou 1. Les lemmes 1.3.2 et 2.4.1 s'appliquent et

$$\begin{aligned} \sum_{n=0}^{\infty} T_\sigma(\mathbf{w}, n) &= \sum_{n=0}^{\infty} (\Delta \mathcal{C}(\mathbf{w}, n) + 2 - \mathcal{P}_\sigma(\mathbf{w}, n) - \mathcal{P}_\sigma(\mathbf{w}, n+1)) \\ &= \sum_{n=0}^{|w|-1} \Delta \mathcal{C}(\mathbf{w}, n) + 2|w| - \sum_{n=0}^{|w|-1} (\mathcal{P}_\sigma(\mathbf{w}, n) + \mathcal{P}_\sigma(\mathbf{w}, n+1)) \\ &= \mathcal{C}(\mathbf{w}, |w|) - \mathcal{C}(\mathbf{w}, 0) + 2|w| - 2 \sum_{n=0}^{|w|-1} \mathcal{P}_\sigma(\mathbf{w}, n) + \mathcal{P}_\sigma(\mathbf{w}, 0) \\ &\quad - \mathcal{P}_\sigma(\mathbf{w}, |w|) \\ &= |w| - 1 + 2|w| - 2 \sum_{n=0}^{|w|-1} \mathcal{P}_\sigma(\mathbf{w}, n) + 1 - \mathcal{P}_\sigma(\mathbf{w}, |w|) \\ &= 3|w| - 2 \sum_{n=0}^{|w|-1} \mathcal{P}_\sigma(\mathbf{w}, n) - \mathcal{P}_\sigma(\mathbf{w}, |w|). \end{aligned}$$

Parallèlement à cela, le théorème 1.3.5 implique que :

$$\begin{aligned}
2\mathcal{D}_\sigma(\mathbf{w}) &= 2\mathcal{D}_\sigma(w^2) \\
&= 2(|w^2| + 1) - 2 \sum_{n=0}^{2|w|} \mathcal{P}_\sigma(w^2, n) \\
&= 4|w| + 2 - 2 \sum_{n=0}^{|w|-1} \mathcal{P}_\sigma(w^2, n) - 2 \sum_{n=|w|}^{2|w|} \mathcal{P}_\sigma(w^2, n) \\
&= 4|w| + 2 - 2 \sum_{n=0}^{|w|-1} \mathcal{P}_\sigma(w^2, n) - |w| - 2 - \mathcal{P}_\sigma(w^2, |w|) \\
&= 3|w| - 2 \sum_{n=0}^{|w|-1} \mathcal{P}_\sigma(w^2, n) - \mathcal{P}_\sigma(w^2, |w|).
\end{aligned}$$

Puisque les facteurs de longueur inférieure ou égale à  $|w|$  sont les mêmes pour  $w^2$  et  $\mathbf{w}$ , on observe que  $2\mathcal{D}_\sigma(\mathbf{w}) = \sum_{n=0}^{\infty} T_\sigma(\mathbf{w}, n)$ , tel qu'annoncé.

Il reste cependant à vérifier le cas où  $\mathbf{w}$  ne contient qu'un nombre fini de  $\sigma$ -palindromes. Cela signifie que la longueur des  $\sigma$ -palindromes est bornée par un entier  $N$ , et il est facile de voir que, pour tout  $n > \min(N, |w|)$ ,

$$\underbrace{\Delta\mathcal{C}(\mathbf{w}, n)}_{=0} + 2 - \underbrace{\mathcal{P}_\sigma(\mathbf{w}, n)}_{=0} - \underbrace{\mathcal{P}_\sigma(\mathbf{w}, n+1)}_{=0} = 2,$$

d'après l'équation (1.1) du lemme 1.3.2. Dans un tel cas, à la fois le  $\sigma$ -défaut et la somme  $\sum_{n \geq 0} T_\sigma(\mathbf{w}, n)$  sont infinis, ce qui conclut la preuve.  $\square$

## 2.5 Mots dont le langage est fermé par $\theta$

On rappelle que  $\theta$  désigne un antimorphisme involutif. La complexité palindromique des mots dont le langage est fermé par miroir est une des premières à avoir

été étudiée. Une des raisons pour cela est que les mots infinis de défaut nul sont récurrents exactement lorsqu'ils sont fermés par miroir (Bucci *et al.*, 2009). Plus généralement, les mots fermés par  $\theta$  possèdent aussi tous la propriété de récurrence, comme l'affirme la proposition suivante, analogue à une proposition citée comme folklore dans (Bucci *et al.*, 2009).

**Proposition 2.5.1.** *Les mots dont le langage est fermé par  $\theta$  sont récurrents.*

*Démonstration.* Soit  $w$  un mot qui n'est pas récurrent. Nous allons montrer qu'il existe certainement un facteur de  $\mathcal{L}(w)$  dont l'image par  $\theta$  n'est pas dans le langage.

Puisque  $w$  n'est pas récurrent, il existe un préfixe  $u$  n'ayant qu'une occurrence dans  $w$  (proposition 0.0.5).

Si  $\hat{u}$  n'est pas facteur de  $w$ , la preuve est terminée. Sinon, soit  $x$  le préfixe de  $w$  terminant par la première occurrence de  $\hat{u}$  et soit  $\alpha$  la lettre suivant le préfixe  $x$  dans  $w$ . Alors,  $\hat{x}\alpha$  n'apparaît pas dans  $w$ . En effet,  $\hat{x}\alpha = \hat{\alpha}\hat{x}$  termine par  $u$  et il y aurait alors une occurrence de  $u$  qui ne soit pas préfixe de  $w$ , contredisant l'unicité de l'occurrence de  $u$ . □

On peut alors en déduire une propriété sur les mots étudiés dans cette section : ils ne sont pas ultimement périodiques, à moins d'être périodiques.

**Corollaire 2.5.2.** *Les mots ultimement périodiques fermés par  $\theta$  sont périodiques.*

*Démonstration.* Si  $w$  est ultimement périodique sans être périodique, il se décompose comme  $w = uv^\omega$ , où  $v$  est primitif et  $u$  n'est pas une puissance entière de  $v$ . Or,  $w$  ne peut être récurrent, car le préfixe  $uv$  n'apparaît qu'une fois. La proposition 2.5.1 implique que son langage n'est pas fermé par  $\theta$ . □





Son langage est fermé par miroir : c'est effectivement un mot récurrent de défaut nul (Glen *et al.*, 2009).

Il n'est cependant pas fermé par  $E$ -miroir ( $E : a \leftrightarrow b$ .) En effet, il n'y a jamais deux occurrences consécutives de la lettre  $a$ , alors qu'il y a des plages arbitrairement longues de la lettre  $b$ .

*Exemple 2.5.5 (Constante de Champernowne, exemple 2.1.4).* La constante de Champernowne, présentée plus haut, a aussi son langage fermé par  $\theta$ , et ce, peu importe l'involution qu'on applique sur les lettres. Ce mot contient en effet tous les facteurs possibles sur l'alphabet des nombres de 0 à 9 et a donc son langage fermé par n'importe quelle transformation  $\{0, \dots, 9\}^+ \rightarrow \{0, \dots, 9\}^+$ .

### Mots uniformément récurrents

*Exemple 2.5.6 (Mots périodiques puissances d'un produit de deux  $\sigma$ -palindromes).* Les mots périodiques sont uniformément récurrents : la distance qui sépare deux occurrences successives d'un même facteur est au plus la période, peu importe la longueur du facteur.

Comme établi par le théorème 1.3.1, les mots périodiques contiennent une infinité de  $\sigma$ -palindromes distincts si et seulement si leur préfixe de la longueur de la période est un produit de deux  $\sigma$ -palindromes. Leur langage est alors fermé par  $\theta$ .

*Exemple 2.5.7 (Mot de Thue-Morse, exemple 0.0.6).* Tel que présenté à la sous-section 1.3.2, les complexités palindromique et  $E$ -palindromique du mot de Thue-Morse ont été étudiées (Blondin Massé *et al.*, 2008b). Parmi les résultats de cet article, on trouve que  $\mathcal{P}_\sigma(\mathbf{t}, n)$  vaut au moins 2 pour tout  $n$  pair (excepté 0), à la fois pour  $\sigma = E$  et pour l'identité (c'est un corollaire de la proposition 7,

(Blondin Massé *et al.*, 2008b)). Comme démontré à la sous-section 1.3.2, c'est un mot dont le langage est fermé par  $\theta$ , que l'involution soit l'échange de lettres ou l'identité.

En outre, il est connu depuis longtemps que le mot de Thue-Morse est uniformément récurrent (Morse et Hedlund, 1938).

*Exemple 2.5.8 (Mots sturmiens)*. Les mots sturmiens (sous-section 1.3.2, puis exemple 2.1) sont aussi des mots uniformément récurrents (Morse et Hedlund, 1940).

Par ailleurs, le comportement des mots sturmiens est complètement différent selon l'involution choisie ( $E$  ou  $\text{Id}$ ), comme le montre la sous-section 1.3.2. En effet, ces mots contiennent soit le facteur  $aa$  ou  $bb$ , mais pas les deux. Ainsi, tout facteur d'un mot sturmien  $s$  contenant  $aa$  (resp.  $bb$ ) n'est pas un  $E$ -palindrome. Comme  $s$  est uniformément récurrent, la longueur des anti-palindromes est bornée et il n'y en a qu'un nombre fini.

En revanche, il est connu que chaque préfixe  $p$  d'un mot sturmien contient  $|p| + 1$   $\text{Id}$ -palindromes (Droubay *et al.*, 2001). Leur langage est fermé par miroir, mais pas par  $E$ -miroir.

## 2.5.2 Théorème principal

Si les derniers énoncés permettent de saisir l'intérêt pour l'étude de la complexité palindromique des mots dont le langage est fermé par  $\theta$ , il reste à montrer le théorème central de cette section :

**Théorème 2.5.9.** *L'équation (2.1) est satisfaite par les mots dont le langage est fermé par  $\theta$ .*

Alors que la preuve du théorème peut sembler simple pour les mots périodiques, celle pour les mots dont le langage est fermé par  $\theta$  est moins intuitive. La preuve du théorème 2.5.9 est largement inspirée par celle connue dans le cas où  $\sigma = \text{Id}$  (Balková *et al.*, 2013). Elle a toutefois été modifiée pour inclure la notion de mots de retours complets, présentée dans les préliminaires.

Comme pour la preuve présentée pour les mots périodiques (théorème 2.4.3), on divise celle-ci en deux scénarios possibles. On suppose d'abord que les deux côtés de l'équation (2.1) sont finis. On montre ensuite que le  $\sigma$ -défaut est infini exactement lorsque la somme  $\sum_{n \geq 0} T_\sigma(\mathbf{w}, n)$  l'est.

### Preuve de l'identité (2.1) dans le cas fini

Nous considérons en premier lieu un mot infini  $\mathbf{w}$  dont le langage est fermé par  $\theta$  et tel que les deux côtés de l'égalité (2.1) sont finis. De tels mots existent, comme en témoignent les exemples 2.5.3, 2.5.4, 2.5.6 et 2.5.8. Étant donné que la situation inverse est présentée à la section 2.4, on peut supposer que  $\mathbf{w}$  est a périodique.

*Remarque 2.5.10.* Étant donné que  $\mathcal{D}_\sigma(\mathbf{w})$  est fini, il existe un entier  $L$  tel que pour toute valeur  $l$  supérieure à  $L$ , le  $\sigma$ -défaut du préfixe de  $\mathbf{w}$  de longueur  $l$  est égal à celui de  $\mathbf{w}$ . Par exemple, pour les mots périodiques, le théorème 1.3.5 implique que  $L \leq |uv| + \lfloor \frac{|u|-|v|}{3} \rfloor$ , où  $uv$  est primitif,  $\mathbf{w} = (uv)^\omega$  et  $u$  et  $v$  sont des  $\sigma$ -palindromes tels que  $|u| \geq |v|$ .

*Remarque 2.5.11.* La finitude de  $\sum_{n=0}^{\infty} T_\sigma(\mathbf{w}, n)$ , de concert avec le lemme 2.1.7 (positivité de  $T_\sigma(\mathbf{w}, n)$ ), implique qu'il existe un entier  $M$  tel que, pour tout  $m$  supérieur à  $M$ ,  $T_\sigma(\mathbf{w}, m) = 0$ .

Les deux dernières remarques nous permettent de définir un entier qui satisfait les deux conditions ci-dessus :  $H = \max\{L, M\} + 1$ . Soit  $q$  le préfixe de  $\mathbf{w}$  de

longueur  $H - 1$ . On remarquera que  $\mathcal{D}_\sigma(q) = \mathcal{D}_\sigma(\mathbf{w})$  et que  $\sum_{n=0}^{\infty} T_\sigma(\mathbf{w}, n) = \sum_{n=0}^{H-1} T_\sigma(\mathbf{w}, n)$ .

Soit  $p$  le plus petit préfixe de  $\mathbf{w}$  qui contienne tous les facteurs de longueur  $H$  de  $\mathbf{w}$ . Bien entendu,  $q$  est un préfixe de  $p$ , ce qui implique que  $\mathcal{D}_\sigma(p) = \mathcal{D}_\sigma(\mathbf{w})$ .

Le théorème 2.3.1 implique alors

$$2\mathcal{D}_\sigma(\mathbf{w}) = 2\mathcal{D}_\sigma(p) = \sum_{n=0}^{|p|} T_\sigma(p, n) = \sum_{n=0}^{H-1} T_\sigma(p, n) + \sum_{n=H}^{|p|} T_\sigma(p, n). \quad (2.2)$$

Par la définition de  $H$ , on sait que  $\sum_{n=H}^{\infty} T_\sigma(\mathbf{w}, n) = 0$ . Le fait que  $p$  et  $\mathbf{w}$  aient le même ensemble de facteurs de longueur  $H$  permet de conclure la preuve à condition que  $\sum_{n=H}^{|p|} T_\sigma(p, n) = 0$ .

Afin de simplifier cette preuve, deux lemmes auxiliaires sont nécessaires :

**Lemme 2.5.12.** *Les définitions plus haut entraînent que*

$$\sum_{n=H}^{|p|} \mathcal{P}_\sigma(p, n) = |p| - H + 1 + \sum_{n=0}^{H-1} (\mathcal{P}_\sigma(q, n) - \mathcal{P}_\sigma(p, n)).$$

*Démonstration.* Considérons l'ensemble des  $\sigma$ -palindromes qui apparaissent dans  $p$  mais pas dans  $q$ . Alors :

$$\#(\text{Pal}_\sigma(p) \setminus \text{Pal}_\sigma(q)) = \sum_{n=0}^{|p|} \mathcal{P}_\sigma(p, n) - \sum_{n=0}^{H-1} \mathcal{P}_\sigma(q, n).$$

Puisque  $\mathcal{D}_\sigma(p) = \mathcal{D}_\sigma(q)$  et  $q$  est un préfixe de  $p$ , la cardinalité de  $\text{Pal}_\sigma(p) \setminus \text{Pal}_\sigma(q)$  doit être au moins  $|p| - H + 1$ . Par la proposition 1.2.1, on obtient directement que  $\#(\text{Pal}_\sigma(p) \setminus \text{Pal}_\sigma(q)) = |p| - H + 1$ .

Pour le reste de la preuve, on procède avec la suite d'égalités :

$$\begin{aligned}
\sum_{n=H}^{|p|} \mathcal{P}_\sigma(p, n) &= \sum_{n=0}^{|p|} \mathcal{P}_\sigma(p, n) - \sum_{n=0}^{H-1} \mathcal{P}_\sigma(p, n) \\
&= |p| - H + 1 + \sum_{n=0}^{H-1} \mathcal{P}_\sigma(q, n) - \sum_{n=0}^{H-1} \mathcal{P}_\sigma(p, n) \\
&= |p| - H + 1 + \sum_{n=0}^{H-1} (\mathcal{P}_\sigma(q, n) - \mathcal{P}_\sigma(p, n)). \quad \square
\end{aligned}$$

Le deuxième lemme porte sur le nombre de facteurs de longueur  $H$  dans  $\mathbf{w}$  qui ne sont pas des  $\sigma$ -palindromes.

**Lemme 2.5.13.** *Soit l'ensemble  $\{u \in \mathcal{L}_H(\mathbf{w}) \mid u \neq \widehat{u}\}$ . Alors, sa cardinalité est*

$$\mathcal{C}(\mathbf{w}, H) - \mathcal{P}_\sigma(\mathbf{w}, H) = 2 \sum_{n=0}^{H-1} (\mathcal{P}_\sigma(p, n) - \mathcal{P}_\sigma(q, n)).$$

*Démonstration.* On procède en créant une bijection entre les paires  $\{u, \widehat{u}\}$  de facteurs distincts et les  $\sigma$ -palindromes de longueur inférieure ou égale à  $H$  dans  $p$ , mais pas dans  $q$ .

— Montrons d'abord que  $\mathcal{C}(\mathbf{w}, H) - \mathcal{P}_\sigma(\mathbf{w}, H) \leq 2 \sum_{n=0}^{H-1} (\mathcal{P}_\sigma(p, n) - \mathcal{P}_\sigma(q, n))$ .

Soit  $u \in \mathcal{L}_H(\mathbf{w})$  qui n'est pas un  $\sigma$ -palindrome. Soit  $s$  le préfixe de  $\mathbf{w}$  qui termine par la première occurrence de  $u$  ou  $\widehat{u}$ . Puisque  $u$  est de longueur  $H$ ,  $|s| \geq H$  et  $\mathcal{D}_\sigma(s) = \mathcal{D}_\sigma(q)$ , ce qui fait que  $\text{PL}\sigma\text{PS}(s)$  est uni-occurent.

Comme  $u$  n'est pas un  $\sigma$ -palindrome, nécessairement,  $\text{PL}\sigma\text{PS}(s) < H$ . Autrement, cela impliquerait  $s = \dots u \dots \widehat{u}$  ou  $s = \dots \widehat{u} \dots u$ , contredisant de fait le choix de  $s$ .

Ainsi, pour toute paire  $\{u, \widehat{u}\}$  de facteurs de longueur  $H$  qui ne sont pas des  $\sigma$ -palindromes, il existe un  $\sigma$ -palindrome de longueur inférieure à

$H$  qui apparaît dans  $p$ , mais pas dans  $q$ ; ce qui donne l'inégalité recherchée.

— Il reste à montrer que

$$2 \sum_{n=0}^{H-1} (\mathcal{P}_\sigma(p, n) - \mathcal{P}_\sigma(q, n)) \leq \mathcal{C}(\mathbf{w}, H) - \mathcal{P}_\sigma(\mathbf{w}, H).$$

Soit  $u$  un  $\sigma$ -palindrome de longueur strictement inférieure à  $H$  qui a au moins une occurrence dans  $p$ , mais aucune dans  $q$ . Soit  $s$  le préfixe de  $\mathbf{w}$  qui termine par la première occurrence de  $u$  ou  $\hat{u}$ . Comme  $\mathcal{D}_\sigma(\mathbf{w}) = \mathcal{D}_\sigma(s)$ , on sait que  $\text{PL}\sigma\text{PS}(s)$  est uni-occurent. Soit  $v$  le suffixe de  $s$  de longueur  $H$ . Alors  $u$  est un suffixe  $\sigma$ -palindromique de  $v$  (différent de  $v$ ), et il découle de la preuve de la proposition 1.2.1 que  $u$  est le seul  $\sigma$ -palindrome uni-occurent de  $s$ . Ainsi,  $v \neq \hat{v}$ .

Comme le langage de  $\mathbf{w}$  est fermé par  $\theta$ , à la fois  $v$  et  $\hat{v}$  font partie de  $\mathcal{L}_H(\mathbf{w}) \setminus \text{Pal}_\sigma(\mathbf{w})$ , ce qui permet de conclure la bijection.  $\square$

**Théorème 2.5.14.** *Soit  $\mathbf{w}$  un mot infini apériodique et fermé par  $\theta$ . Si  $\mathcal{D}_\sigma(\mathbf{w})$  est fini et si  $\sum_{n=0}^{\infty} T_\sigma(\mathbf{w}, n)$  l'est, l'équation (2.1) est satisfaite.*

*Démonstration.* Pour achever cette preuve, il suffit de montrer que  $\sum_{n=H}^{|\mathbf{p}|} T_\sigma(p, n)$  est nul. Ensuite, on complète directement à partir de l'équation (2.2) :

$$\begin{aligned}
\sum_{n=H}^{|p|} T_{\sigma}(p, n) &= \sum_{n=H}^{|p|} (\Delta \mathcal{C}(p, n) + 2 - \mathcal{P}_{\sigma}(p, n+1) - \mathcal{P}_{\sigma}(p, n)) \\
&= \underbrace{\mathcal{C}(p, |p|+1)}_{=0} - \mathcal{C}(p, H) + 2(|p| - H + 1) - 2 \sum_{n=H}^{|p|} \mathcal{P}_{\sigma}(p, n) \\
&\quad - \underbrace{\mathcal{P}_{\sigma}(p, |p|+1)}_{=0} + \mathcal{P}_{\sigma}(p, H) \\
&= \mathcal{P}_{\sigma}(p, H) - \mathcal{C}(p, H) + 2(|p| - H + 1) - 2 \sum_{n=H}^{|p|} \mathcal{P}_{\sigma}(p, n) \\
&= \mathcal{P}_{\sigma}(p, H) - \mathcal{C}(p, H) + 2(|p| - H + 1) - 2(|p| - H + 1 \\
&\quad - \sum_{n=0}^{H-1} (\mathcal{P}_{\sigma}(p, n) - \mathcal{P}_{\sigma}(q, n))) \tag{2.3}
\end{aligned}$$

$$\begin{aligned}
&= -2 \sum_{n=0}^{H-1} (\mathcal{P}_{\sigma}(p, n) - \mathcal{P}_{\sigma}(q, n)) + 2 \sum_{n=0}^{H-1} (\mathcal{P}_{\sigma}(p, n) - \mathcal{P}_{\sigma}(q, n)) \tag{2.4} \\
&= 0.
\end{aligned}$$

Il est de mise de noter que (2.3) est obtenue par le lemme 2.5.12 et (2.4) découle du lemme 2.5.13.  $\square$

### Preuve de l'identité (2.1) dans le cas infini

Supposons maintenant que  $\sum_{n=0}^{\infty} T_{\sigma}(\mathbf{w}, n)$  est infini. On doit alors montrer que le  $\sigma$ -défaut l'est aussi. Réciproquement, on doit démontrer que lorsque le  $\sigma$ -défaut est infini, la somme des  $T_{\sigma}(\mathbf{w}, n)$ ,  $n \in \mathbb{N}$ , est aussi infinie. Pour y parvenir, nous allons utiliser les notions de mots de retour complet et de graphes de Rauzy, introduites dans les préliminaires, puis donner quelques lemmes qui mèneront à la preuve. Cette approche n'est pas tout à fait originale, puisque les auteures et auteurs de (Baláži *et al.*, 2007; Starosta, 2011) ont utilisé les graphes de Rauzy

pour présenter une preuve du lemme 2.1.7, puis, plus tard, pour démontrer l'identité de Brlek-Reutenauer dans le cas où  $\sigma = \text{Id}$ .

Comme énoncé dans les préliminaires, l'ensemble des mots de retour complet de n'importe quel facteur dans un mot infini est bien défini exactement lorsque le mot est récurrent. Par la proposition 2.5.1, c'est le cas des mots étudiés dans cette section. De plus, on suppose que tous les mots sont apériodiques; comme démontré à la section 2.4, les mots périodiques satisfont l'équation (2.1).

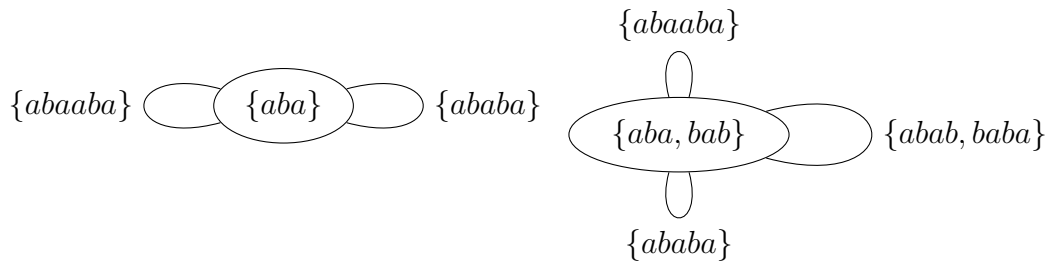
Pour la preuve générale du théorème central de cette section (théorème 2.5.9), on doit construire une version légèrement différente du graphe de Rauzy réduit. On utilisera ainsi le fait que le langage des mots étudiés est fermé par  $\theta$ .

Soit  $G_\sigma(\mathbf{w}, n)$  le graphe non-orienté ayant comme sommets les paires  $\{u, \hat{u}\}$ , pour tout  $u \in \mathcal{L}_n(\mathbf{w})$  qui est un facteur spécial (gauche ou droit). Il existe une arête entre deux sommets  $\{u, \hat{u}\}$  et  $\{v, \hat{v}\}$  exactement lorsqu'il existe un entier positif  $m$  tel qu'il y a dans  $\mathcal{L}_{n+m}(\mathbf{w})$  un facteur spécial (gauche ou droit)  $s$  qui débute soit par  $u$  ou  $\hat{u}$  et termine par  $v$  ou  $\hat{v}$ . En conséquence, toute arête correspond à une paire de facteurs  $\{s, \hat{s}\}$  de  $\mathcal{L}_{n+m}(\mathbf{w})$  et il peut y avoir plusieurs arêtes entre deux sommets.

Les étiquettes d'une chaîne de  $\{u, \hat{u}\}$  à  $\{v, \hat{v}\}$  correspondent à un facteur de  $\mathbf{w}$  ayant pour préfixe  $u$  ou  $\hat{u}$  et pour suffixe  $v$  ou  $\hat{v}$ . Les lettres entre le préfixe et le suffixe sont données par les étiquettes des arêtes (on écrit les facteurs spéciaux une seule fois).

**Lemme 2.5.15.** *Soit  $\mathbf{w}$  un mot infini apériodique dont le langage est fermé par  $\theta$ .*





**Figure 2.1** Graphes  $G_{\text{Id}}(\text{Fib}, 3)$  (à gauche) et  $G_E(\text{Fib}, 3)$  (à droite) du mot de Fibonacci. La lectrice ou le lecteur alerte remarquera que le graphe de gauche a toutes ses arêtes et sommets étiquetés par des palindromes, alors que le même énoncé n'est pas vrai pour le graphe de droite. Une explication est notamment que les mots sturmiens, comme le mot de Fibonacci, satisfont  $\mathcal{D}_{\text{Id}}(\mathbf{s}) = 0$  et  $\mathcal{D}_E(\mathbf{s}) = \infty$ , tout comme ils vérifient l'équation (2.1).

Alors,

1. Pour une chaîne simple (orientée) et un facteur initial donnés dans  $G_\sigma(\mathbf{w}, n)$ , il existe un unique facteur correspondant aux étiquettes de la chaîne.
2. Pour une chaîne simple donnée dans  $G_\sigma(\mathbf{w}, n)$ , il existe au plus deux mots correspondant aux étiquettes de la chaîne.

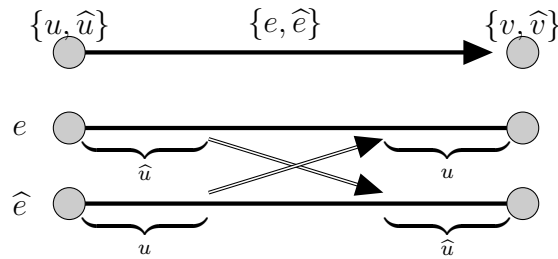
*Démonstration.* Suivant les hypothèses du lemme, on montre les énoncés dans le même ordre :

1. La preuve se fait par récurrence sur la longueur de la chaîne, c'est-à-dire son nombre d'arêtes. Considérons la chaîne simple  $p$  de  $G_\sigma(\mathbf{w}, n)$  qui part du sommet  $\{u, \hat{u}\}$  et termine au sommet  $\{v, \hat{v}\}$ .

Si  $p$  est de longueur 1, l'unique arête  $e$  de la chaîne correspond à un ou

deux facteurs de  $\mathbf{w}$ . S'il y en a deux, chacun est l'image par  $\theta$  de l'autre (par construction). Notons ces facteurs  $x$  et  $\hat{x}$ .

Considérons le mot,  $x$  ou  $\hat{x}$ , qui débute par  $u$ . Pour voir que ce mot est unique, on suppose au contraire que  $x$  et  $\hat{x}$  débutent par  $u$  et on remarque que le fait qu'ils soient réciproquement l'image par  $\theta$  de l'autre force que l'arête soit une boucle. Or, une boucle n'est pas une chaîne simple, car elle passe deux fois par le même sommet.

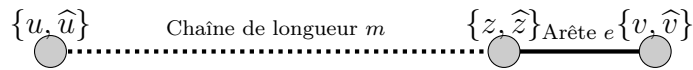


**Figure 2.2** Pour un facteur de longueur  $n$  et une arête du graphe  $G_\sigma(\mathbf{w}, n)$  donnés, il existe un unique facteur de  $\mathbf{w}$  correspondant à l'étiquette.

Pour la suite, on suppose que, pour toute chaîne de longueur inférieure ou égale à  $m$ , il existe un unique mot correspondant aux étiquettes de la chaîne (orientée) et débutant par un facteur donné.

Si  $p$  est une chaîne de longueur  $m + 1$ , elle peut s'exprimer comme la concaténation d'une chaîne de longueur  $m$  de  $\{u, \hat{u}\}$  à  $\{z, \hat{z}\}$  et d'une arête  $e$  de  $\{z, \hat{z}\}$  à  $\{v, \hat{v}\}$ .

Par hypothèse de récurrence, il existe un seul mot  $y$  qui corresponde aux étiquettes de la chaîne de  $\{u, \hat{u}\}$  à  $\{z, \hat{z}\}$  et qui a pour préfixe  $u$ .



**Figure 2.3** Une chaîne de longueur  $m + 1$  dans  $G_\sigma(\mathbf{w}, n)$ .

Comme les  $n$  dernières lettres de  $y$  sont aussi les  $n$  premières du mot associé à l'arête  $e$ , on utilise le cas de base et on trouve qu'il existe un seul mot possible pour étendre  $y$  en suivant  $p$ . C'est donc le seul mot possible commençant par  $u$  et suivant la chaîne  $p$ .

2. Le deuxième énoncé est une conséquence du premier. En effet, chaque sommet de  $G_\sigma(\mathbf{w}, n)$  correspond à au plus deux facteurs de longueur  $n$  dans  $\mathbf{w}$ . Pour chacun de ces facteurs, il existe un unique mot correspondant à une chaîne donnée; c'est le premier énoncé.  $\square$

**Si  $\sum_{n=0}^{\infty} T_\sigma(\mathbf{w}, n)$  est fini :**

À l'aide des graphes de Rauzy et de leurs variantes, Štěpán Starosta a montré deux lemmes, dont un que nous exprimons en termes de mots de retour complet.

**Lemme 2.5.16** (Corollaire 7, (Starosta, 2011)). *Soit  $\mathbf{w}$  un mot infini dont le langage est fermé par  $\theta$  et soit  $n > 0$ . Alors,  $T_\sigma(\mathbf{w}, n) = 0$  si et seulement si les deux conditions suivantes sont rencontrées :*

1. *Le graphe obtenu de  $G_\sigma(\mathbf{w}, n)$  en enlevant les boucles – les arêtes d'un sommet vers lui-même – est un arbre.*
2. *L'étiquette associée à toute boucle dans  $G_\sigma(\mathbf{w}, n)$  est un singleton formé d'un  $\sigma$ -palindrome.*

**Lemme 2.5.17** (Analogie au lemme 14, (Starosta, 2011)). *Soit  $\mathbf{w}$  un mot infini dont le langage est fermé par  $\theta$ . Si  $T_\sigma(\mathbf{w}, n) = 0$  pour un certain entier  $n$ , alors pour tout  $u \in \mathcal{L}_n(\mathbf{w})$ , les occurrences de  $u$  et de  $\hat{u}$  s'alternent.*

Ce dernier lemme est exprimé dans l'article original (Starosta, 2011) pour le cas particulier  $n = 1$ , mais la preuve est donnée pour toute valeur de  $n$ .

On peut traduire le lemme 2.5.16 en termes de mots de retour complet :

**Lemme 2.5.18** (Conséquence du lemme 2.5.16). *Soit  $\mathbf{w}$  un mot infini aperiodique dont le langage est fermé par  $\theta$  et soit  $n > 0$ . Alors,  $T_\sigma(\mathbf{w}, n) = 0$  si et seulement si les deux conditions suivantes sont satisfaites :*

1. *Étant donné deux paires distinctes  $\{u, \hat{u}\}$  et  $\{v, \hat{v}\}$  de facteurs spéciaux de  $\mathcal{L}_n(\mathbf{w})$ , il existe un unique facteur de  $\mathbf{w}$  débutant par  $u$  et terminant par  $v$  ou  $\hat{v}$ , sans autre occurrence de  $u$ ,  $\hat{u}$ ,  $v$  ou  $\hat{v}$  et qui ne contienne plus d'une occurrence de la paire  $\{z, \hat{z}\}$ , où  $z$  est un facteur spécial de longueur  $n$  dans  $\mathbf{w}$ .*
2. *Soit  $u \in \mathcal{L}_n(\mathbf{w})$  un facteur spécial. Tout mot de  $\theta$ -retour complet de  $u$  qui ne contient pas de facteur spécial de longueur  $n$ , autre que  $u$  et  $\hat{u}$ , est un  $\sigma$ -palindrome.*

*Démonstration.* Comme les hypothèses des deux lemmes sont identiques, il est suffisant de montrer que les énoncés 1 et 2 du lemme 2.5.16 sont équivalents à ceux du lemme 2.5.18.

1. Un graphe est un arbre si et seulement si chaque paire de sommets est reliée par une unique chaîne (proposition 0.0.20).

Une chaîne dans un graphe passe au plus une fois par chacun des sommets.

Dans  $G_\sigma(\mathbf{w}, n)$ , cela veut dire que pour chaque paire  $\{z, \hat{z}\}$  de facteurs de

$\mathcal{L}_n(\mathbf{w})$ , il y a au plus une occurrence d'un de ces facteurs (et donc au plus un des deux apparaît).

Le lemme 2.5.15 implique que pour une chaîne donnée et un préfixe  $u$ , il n'y a qu'une seule possibilité pour les étiquettes des arêtes. Donc le graphe est un arbre si et seulement s'il existe un unique facteur de  $\mathbf{w}$  qui commence par  $u$  et termine par  $v$  ou  $\widehat{v}$  sans qu'il n'y ait plus d'une occurrence d'un facteur de longueur  $n$  ou de son image par  $\theta$ .

2. Par définition du graphe  $G_\sigma(\mathbf{w}, n)$ , les boucles correspondent à des mots de  $\theta$ -retour complet sans facteur spécial de longueur  $n$  autre que le préfixe et le suffixe. □

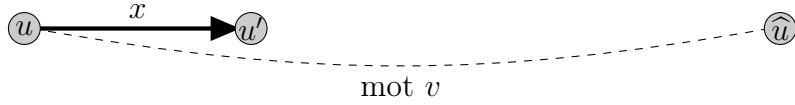
**Théorème 2.5.19.** *Soit  $\mathbf{w}$  un mot infini apériodique dont le langage est fermé par  $\theta$ . Supposons qu'il existe un entier  $N$  tel que pour tout  $n \geq N$ ,  $T_\sigma(\mathbf{w}, n) = 0$ . Alors, quelque soit  $u \in \mathcal{L}_n(\mathbf{w})$ ,  $n \geq N$ , tout mot de  $\theta$ -retour complet  $v$  de  $u$  est un  $\sigma$ -palindrome.*

*Démonstration.* On remarque d'abord que les hypothèses du lemme 2.5.17 sont satisfaites. Ainsi,  $v$  commence par  $u$  et termine par  $\widehat{u}$ .

Le reste de la preuve se fait en deux cas.

- Si  $u$  est un facteur spécial droit : il y a encore deux cas possibles. Si  $v$  ne contient pas d'autre facteur spécial de longueur  $n$ , l'énoncé 2 du lemme 2.5.18 s'applique et  $v$  est un  $\sigma$ -palindrome.

Sinon, on note  $u'$  le premier facteur spécial droit de longueur  $n$  à droite de  $u$ . Soit  $x$  le facteur commençant par  $u$  et terminant par  $u'$ , sans autre occurrence de  $u$ ,  $u'$ ,  $\widehat{u}$  ou  $\widehat{u}'$ .



**Figure 2.4** Récurrence sur les facteurs spéciaux droits du mot  $v$ .

Puisque  $|x| > N$  et que  $x$  est un facteur spécial de  $\mathbf{w}$ ,  $x$  satisfait les hypothèses du lemme 2.5.18 et on applique le même argument sur les facteurs de longueur  $|x|$ . Si  $v$  ne contient pas d'autre facteur spécial de longueur  $|x|$ , on utilise le premier cas. Sinon, on répète la procédure utilisée pour  $u$  en choisissant  $x'$  comme le facteur de  $v$  terminant par le facteur spécial suivant  $x$ . On peut appliquer cette procédure récursivement jusqu'à ce qu'il n'y ait plus de grand facteur spécial autre que le préfixe et le suffixe.

- Si  $u$  n'est pas un facteur spécial droit : soit  $v'$  le plus court préfixe de  $v$  qui est un facteur spécial droit et qui est plus long que  $u$ . Nécessairement, il existe un tel facteur, car  $\mathbf{w}$  n'est pas ultimement périodique (lemme 2.5.2). Ainsi,  $v$  est aussi le mot de  $\theta$ -retour de  $v'$ . Comme  $|v'| > |u|$ ,  $T_\sigma(\mathbf{w}, |v'|) = 0$  et on peut se ramener au premier cas.  $\square$

On a maintenant tous les ingrédients pour prouver que le  $\sigma$ -défaut est fini. C'est l'objet du prochain théorème.

**Théorème 2.5.20.** *Soit  $\mathbf{w}$  un mot infini aperiodique dont le langage est fermé par  $\theta$ . Si  $\sum_{n=0}^{\infty} T_\sigma(\mathbf{w}, n) < \infty$ , alors le  $\sigma$ -défaut de  $\mathbf{w}$  est fini.*

*Démonstration.* Si  $\sum_{n=0}^{\infty} T_\sigma(\mathbf{w}, n) < \infty$ , la remarque 2.5.11 indique qu'il existe un entier  $M$  tel que  $T_\sigma(\mathbf{w}, m) = 0$ , pour tout  $m \geq M$ . Les hypothèses du théorème 2.5.19 sont alors satisfaites et les mots de  $\theta$ -retour complet de facteurs de longueur

$m \geq M$  sont des  $\sigma$ -palindromes.

Pour montrer que le  $\sigma$ -défaut est fini, on observe d'abord que le  $\sigma$ -défaut est fini si et seulement si, à partir d'un certain rang  $L$ , le plus long  $\sigma$ -palindrome suffixe de tout préfixe de  $\mathbf{w}$  de longueur  $l \geq L$  est uni-occurent (remarque 2.5.10).

On suppose que le préfixe  $p$  de  $\mathbf{w}$  de longueur  $m \geq M$  est tel que  $\text{PL}\sigma\text{PS}(p)$  soit apparu à une position antérieure. Puisque  $m > M$ , le mot qui débute à l'avant-dernière occurrence de  $\text{PL}\sigma\text{PS}(p)$  et qui est suffixe de  $p$  est un mot de  $\theta$ -retour complet et un  $\sigma$ -palindrome, ce qui contredit la maximalité de  $\text{PL}\sigma\text{PS}(p)$ . D'où  $\text{PL}\sigma\text{PS}(p)$  doit être uni-occurent et le défaut de  $\mathbf{w}$ , nécessairement fini.  $\square$

Le dernier théorème achève de démontrer que si la somme  $\sum_{n=0}^{\infty} T_{\sigma}(\mathbf{w}, n)$  est finie, le défaut l'est aussi. Il reste toutefois à montrer la réciproque.

**Si le  $\sigma$ -défaut est fini :**

Pour ce côté de la preuve, on procède dans le sens contraire : on décrit d'abord ce qu'un  $\sigma$ -défaut fini implique sur les mots de  $\theta$ -retour complet, puis on termine avec les conséquences de cela sur les valeurs de  $T_{\sigma}(\mathbf{w}, n)$ .

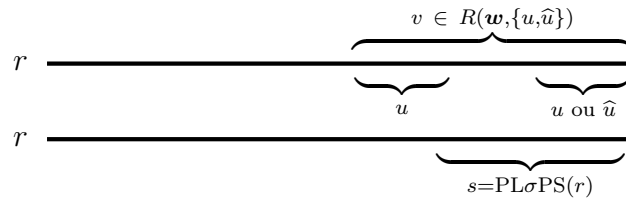
**Théorème 2.5.21.** *Soit  $\mathbf{w}$  un mot infini aperiodique dont le langage est fermé par  $\theta$  et dont le  $\sigma$ -défaut est fini. Alors il existe un entier  $N$  tel que, pour tout facteur  $u$  de longueur supérieure à  $N$ , tout mot de  $\theta$ -retour complet de  $u$  est un  $\sigma$ -palindrome.*

*Démonstration.* On procède par contradiction.

D'après la remarque 2.5.10, un  $\sigma$ -défaut fini implique qu'il existe un entier  $L$  tel que, pour tout préfixe  $p$  de longueur supérieure à  $L$  dans  $\mathbf{w}$ ,  $\text{PL}\sigma\text{PS}(p)$  est uni-

occurrent. On pose  $N$  égal à  $L$  et  $u$  tel que  $|u| > N$ . Soit  $v$  un mot de  $\theta$ -retour complet de  $u$  dans  $\mathbf{w}$ . Supposons qu'il n'est pas un  $\sigma$ -palindrome.

Sans perte de généralité, on peut supposer que  $v$  est le premier tel facteur qui apparaît dans  $\mathbf{w}$ . On note  $r$  le préfixe de  $\mathbf{w}$  qui termine par la première occurrence de  $v$  et soit  $s = \text{PL}\sigma\text{PS}(r)$ . Par hypothèse,  $s$  doit être uni-occurrent dans  $r$ .



**Figure 2.5** Préfixe de  $\mathbf{w}$  terminant par le premier mot de  $\theta$ -retour complet de  $u$ .

Ainsi :

1. Si  $|s| \leq |u|$ ,  $s$  n'est pas uni-occurrent puisqu'il y a deux occurrences de  $u$  ou  $\hat{u}$  et  $s$  est un  $\sigma$ -palindrome.
2. Si  $|u| < |s| < |v|$ , alors  $v$  contient au moins trois occurrences de la paire  $\{u, \hat{u}\}$  (car  $s$  est un  $\sigma$ -palindrome), ce qui contredit l'hypothèse comme quoi  $v$  est un mot de  $\theta$ -retour de  $u$ .
3. L'égalité  $s = v$  est impossible, puisque cela contredirait le fait que  $s$  est un  $\sigma$ -palindrome, alors que  $v$  n'en est pas un.
4. Si  $|v| < |s|$ , il existe au moins deux occurrences de la paire  $\{v, \hat{v}\}$  dans  $s$  (car c'est un  $\sigma$ -palindrome), ce qui contredit le choix de  $r$ .

On en conclut donc que  $v$  est un  $\sigma$ -palindrome. □

Afin de conclure que la somme  $\sum_{n=0}^{\infty} T_{\sigma}(\mathbf{w}, n)$  est finie lorsque le défaut l'est, on utilise le dernier théorème ainsi que le lemme 2.5.18. C'est ce qu'explique ce



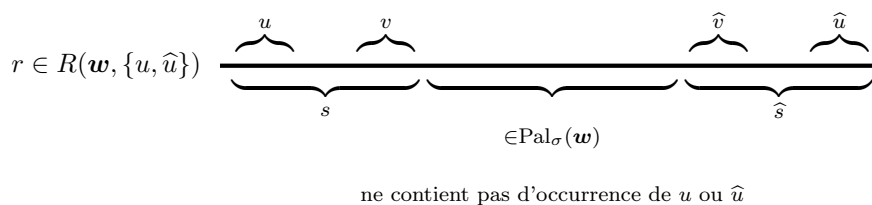
qui suit :

**Théorème 2.5.22.** *Soit  $\mathbf{w}$  un mot apériodique dont le langage est fermé par  $\theta$ . Si le  $\sigma$ -défaut est fini, alors  $\sum_{n=0}^{\infty} T_{\sigma}(\mathbf{w}, n)$  est également fini.*

*Démonstration.* Comme  $\mathcal{D}_{\sigma}(\mathbf{w})$  est fini, on peut appliquer le théorème 2.5.21. On sait donc que, si  $u$  est un facteur suffisamment long de  $\mathbf{w}$ , tout mot de  $\theta$ -retour complet de  $u$  est un  $\sigma$ -palindrome. Pour montrer que  $\sum_{n=0}^{\infty} T_{\sigma}(\mathbf{w}, n) < \infty$  à partir du lemme 2.5.18, on cherche à montrer que  $G_{\sigma}(\mathbf{w}, n)$  satisfait les deux conditions suivantes lorsque  $n$  est assez grand :

1. *Étant donné deux paires distinctes  $\{u, \hat{u}\}$  et  $\{v, \hat{v}\}$  de facteurs spéciaux de  $\mathcal{L}_n(\mathbf{w})$ , il existe un unique facteur de  $\mathbf{w}$  débutant par  $u$  et terminant par  $v$  ou  $\hat{v}$ , sans autre occurrence de  $u, \hat{u}, v$  ou  $\hat{v}$  et qui ne contienne plus d'une occurrence de la paire  $\{z, \hat{z}\}$ , où  $z$  est un facteur spécial de longueur  $n$  dans  $\mathbf{w}$ .*

Soit  $s$  un facteur de  $\mathbf{w}$  débutant par une occurrence de  $u$  et terminant par  $v$  sans autre occurrence de  $u, \hat{u}, v$  ou  $\hat{v}$ . Soit  $r$  un facteur de  $\mathbf{w}$  ayant pour préfixe  $s$  et qui est un mot de  $\theta$ -retour complet de  $u$ . Par le théorème 2.5.21,  $r$  est un  $\sigma$ -palindrome et  $\hat{s}$  est un suffixe de  $r$ . De plus,  $\hat{s}$  est le premier facteur suivant  $s$  qui ait pour préfixe  $u$  ou  $\hat{u}$  et pour suffixe  $v$  ou  $\hat{v}$ , ou le contraire, sans autre occurrence de  $u$  ou de  $\hat{u}$  (parce que  $r \in R(\mathbf{w}, \{u, \hat{u}\})$ ).



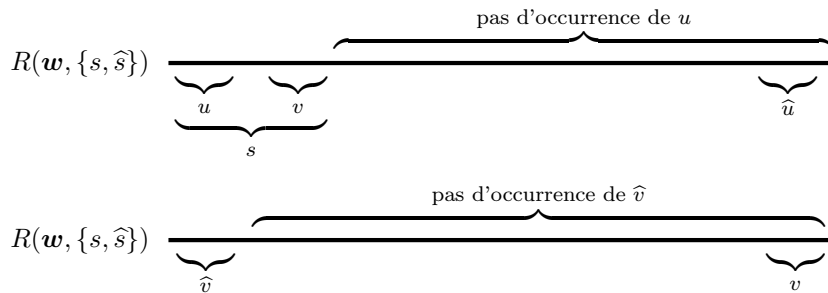
**Figure 2.6** Structure de  $r$ , un mot de  $\theta$ -retour complet de  $u$  dans  $\mathbf{w}$ .

Le mot de  $\theta$ -retour de  $\widehat{v}$  (le préfixe de  $\widehat{s}$ ) est aussi un  $\sigma$ -palindrome, car  $|\widehat{v}| = |u|$ , et il termine par  $\widehat{s} = s$ .

Ceci achève de montrer qu'il n'y a qu'une possibilité pour  $s$ , en tant que facteur de  $\mathbf{w}$  débutant par une occurrence de  $u$  et terminant par  $v$  sans autre occurrence de  $u$ ,  $\widehat{u}$ ,  $v$  ou  $\widehat{v}$ .

De plus, il n'y a pas de facteur  $x$  de  $\mathbf{w}$  débutant par  $u$  et terminant par  $\widehat{v} \neq v$  sans autre occurrence de  $u$ ,  $\widehat{u}$ ,  $v$  ou  $\widehat{v}$ . En effet, on remarque qu'à partir de la première occurrence de  $u$ , tout ce qui suit est soit partie d'un mot de  $\theta$ -retour complet de  $s$  ou de  $\widehat{s}$ .

- (a)  $x$  ne peut faire partie d'un mot de  $\theta$ -retour de  $s$  : Les mots de  $\theta$ -retour complet de  $s$  sont aussi des mots de  $\theta$ -retour complet de  $u$  (parce que les mots de  $\theta$ -retour complet de  $u$  sont des  $\sigma$ -palindromes). Ainsi, ils débutent par  $u$  et terminent par  $\widehat{u}$  sans autre occurrence de  $u$  ou  $\widehat{u}$ . La première occurrence d'un facteur de la paire  $\{v, \widehat{v}\}$  en est une de  $v$ ; c'est le suffixe de  $s$ . Par conséquent,  $x$  n'est pas facteur d'un mot de  $\theta$ -retour complet de  $s$ .



**Figure 2.7** Mots de  $\theta$ -retour complet de  $s$  et de  $\widehat{s}$  dans  $\mathbf{w}$ .

- (b)  $x$  ne peut faire partie d'un mot de  $\theta$ -retour de  $\widehat{s}$  : On procède de la

même façon en remarquant que les mots de  $\theta$ -retour de  $\hat{s}$  sont aussi des mots de  $\theta$ -retour de  $\hat{v}$ . Ils débutent alors par  $\hat{v}$  et n'ont pas d'autre occurrence de  $\hat{v}$ , car les mots de  $\theta$ -retour complet de  $\hat{v}$  sont des palindromes et qu'ils terminent par  $v$ . Autrement dit,  $x = u \dots \hat{v}$  ne peut être facteur d'un mot de  $\theta$ -retour de  $\hat{s}$ .

2. Soit  $u \in \mathcal{L}_n(\mathbf{w})$  un facteur spécial. Tout mot de  $\theta$ -retour complet de  $u$  qui ne contient pas de facteur spécial de longueur  $n$ , autre que  $u$  et  $\hat{u}$ , est un  $\sigma$ -palindrome.

Comme tout mot de  $\theta$ -retour complet de  $u$  est un  $\sigma$ -palindrome (théorème 2.5.21), cette condition est satisfaite.  $\square$

*Démonstration du théorème 2.5.9.* C'est une conséquence directe des théorèmes 2.4.3, 2.5.14, 2.5.20 et 2.5.22.  $\square$

## 2.6 Mots uniformément récurrents

**Théorème 2.6.1.** *L'équation (2.1) est satisfaite par les mots uniformément récurrents.*

Une preuve de ce théorème a déjà été donnée dans le cas spécifique où  $\sigma = \text{Id}$  (Balková *et al.*, 2011b; Balková *et al.*, 2011a). Cependant, elle pourrait être considérablement simplifiée en utilisant le résultat de la section 2.5. La preuve se découpe en deux parties. Soit le résultat de la section 2.5 s'applique, soit le défaut et  $\sum_{n \geq 0} T_\sigma(\mathbf{w}, n)$  sont infinis. Elle est donnée après ces deux lemmes.

**Lemme 2.6.2.** *L'équation (2.1) est satisfaite pour tout mot contenant un nombre fini de  $\sigma$ -palindromes.*

*Démonstration.* Soit  $w$  un mot contenant un nombre fini de  $\sigma$ -palindromes. Si  $w$  est de longueur finie, on sait qu'il satisfait l'équation (2.1) par le théorème 2.3.1. Si  $w$  est infini, il est nécessairement de défaut infini. De plus, pour tout  $n$  à partir d'un certain rang  $N$ ,  $P_\sigma(w, n) = 0$  et donc

$$T_\sigma(w, n) = \Delta\mathcal{C}(w, n) + 2 \geq 2,$$

car  $\Delta\mathcal{C}(w, n)$  est toujours positive (proposition 0.0.2). En particulier, la somme  $\sum_{n=0}^{\infty} T_\sigma(w, n)$  est infinie, tout comme le défaut.  $\square$

**Lemme 2.6.3.** *Soit  $\mathbf{w}$  un mot infini uniformément récurrent. Si son langage n'est pas fermé par  $\theta$ ,  $\mathbf{w}$  contient un nombre fini de  $\sigma$ -palindromes.*

*Démonstration.* Soit  $u$  un facteur de  $\mathbf{w}$  tel que  $\hat{u} \notin \mathcal{L}(\mathbf{w})$ .

Comme  $\mathbf{w}$  est uniformément récurrent, il existe un entier  $m$  tel que deux occurrences consécutives de  $u$  sont toujours à distance au plus  $m$  l'une de l'autre. Il n'existe donc aucun facteur de  $\mathbf{w}$  de longueur supérieure à  $m$  qui soit un palindrome.

Comme l'alphabet est fini, le langage  $\sigma$ -palindromique l'est aussi.  $\square$

Cela nous mène maintenant à la preuve du théorème 2.6.1 :

*Démonstration du théorème 2.6.1.* Soit  $\mathbf{w}$  un mot infini uniformément récurrent. Si le langage de  $\mathbf{w}$  est fermé par  $\theta$ , on sait par le théorème 2.5.9 que l'équation

(2.1) est satisfaite. Autrement,  $\mathbf{w}$  satisfait les hypothèses du lemme 2.6.3 et ne contient qu'un nombre fini de palindromes. D'où l'équation (2.1) est vérifiée.  $\square$

## 2.7 Points fixes de morphismes primitifs

**Théorème 2.7.1.** *L'équation (2.1) est satisfaite pour les points fixes de morphismes primitifs.*

La preuve de ce théorème repose sur un lemme déjà connu. Nommé « lemme de la fenêtre » par ses auteurs, il affirme que tout point fixe de morphisme primitif est uniformément récurrent et que la distance entre deux occurrences consécutives de n'importe quel facteur est proportionnelle à sa longueur :

**Lemme 2.7.2** (Lemme 2.1, (Damanik et Zare, 2000)). *Soit  $\Psi$  un morphisme primitif. Alors il existe une constante finie  $C_\Psi$  telle que dans tout point fixe infini  $\mathbf{w}$  de  $\Psi$ , tout facteur  $u$  ayant une occurrence dans  $\mathbf{w}$  est facteur de tous les sous-mots de longueur  $C_\Psi \cdot |u|$ .*

Autrement dit, les occurrences de tout facteur de  $\mathbf{w}$  apparaissent à distance bornée linéairement (en fonction de la longueur du facteur) l'une de l'autre et le mot est par conséquent uniformément récurrent.

*Preuve du théorème 2.7.1.* En utilisant le lemme de la fenêtre (lemme 2.7.2) et le théorème 2.6.1, on obtient que, puisqu'ils sont uniformément récurrents, ils satisfont l'identité.  $\square$

## 2.8 Une identité vraie pour les mots récurrents?

Les notions des dernières sections montrent que plusieurs types de mots récurrents satisfont l'identité. Parmi celles-ci, deux grandes familles distinctes englobent d'autres : les mots uniformément récurrents et les mots dont le langage est fermé par  $\theta$ .

### 2.8.1 Mots ayant un nombre fini de palindromes

Il existe aussi des mots récurrents satisfaisant l'identité en question qui n'appartiennent à aucune de ces deux familles. Parmi ceux-ci, tous les mots ayant un nombre fini de palindromes satisfont l'identité, par le lemme 2.6.2.

Lorsque  $\sigma = \text{Id}$ , on retrouve notamment les suites obtenues par pliage de papier, comme la courbe du Dragon (A014707 de l'OEIS). Cette suite est construite récursivement en prenant la limite  $n \rightarrow \infty$  de

$$w_0 = \varepsilon, \quad w_{n+1} = w_n 0 \widetilde{w}_n.$$

Elle est dite suite par pliage de papier régulière parce qu'elle correspond au résultat (vu de côté) d'un papier plié en deux chaque fois dans le même sens.

Comme le faisaient déjà remarquer Srećko Brlek et Christophe Reutenauer (Brlek et Reutenauer, 2011), elles satisfont l'équation puisqu'elles ne contiennent aucun palindrome de longueur supérieure à 14, un résultat dû à Jean-Paul Allouche (Allouche, 1997).

Il faut cependant noter que les mots ayant un nombre fini de palindromes ne sont pas nécessairement récurrents.

### 2.8.2 Un problème ouvert

Ce chapitre a présenté plusieurs exemples de mots récurrents satisfaisant l'équation (2.1). Il est connu depuis l'énonciation de la conjecture que la condition de récurrence est nécessaire. Le contre-exemple suivant démontre que l'égalité de l'équation (2.1) n'est pas toujours satisfaite par les mots qui ne sont pas récurrents :

*Exemple 2.8.1* (Similaire à un contre-exemple de (Brlek et Reutenauer, 2011)).  
Considérons le mot de Fibonacci (sur l'alphabet binaire  $\{a, b\}$ ).

Il est facile de voir que le mot infini  $c \cdot \text{Fib}$  n'est pas récurrent lorsque  $\text{Fib}$  est le mot de Fibonacci : il ne contient qu'une occurrence de la lettre  $c$ .

En outre, il ne satisfait pas l'équation (2.1). Son défaut est le même que celui du mot de Fibonacci : puisque, pour chaque préfixe  $p$  de  $\text{Fib}$ ,  $\text{PLPS}(p)$  est uni-occurent dans  $p$ , l'ajout d'une lettre n'apparaissant pas dans  $p$  n'altère pas cette propriété (et  $\text{PLPS}(c)$  est uni-occurent).

En revanche,

$$\sum_{n \geq 0} T_{\text{Id}}(c \cdot \text{Fib}, n) = -1 + \sum_{n \geq 0} T_{\text{Id}}(\text{Fib}, n).$$

On obtient cela directement par calculs après avoir remarqué que

$$\mathcal{C}(c \cdot \text{Fib}, n) = \begin{cases} \mathcal{C}(\text{Fib}, n) & \text{si } n = 0; \\ \mathcal{C}(\text{Fib}, n) + 1 & \text{sinon.} \end{cases}$$

et

$$\mathcal{P}(c \cdot \text{Fib}, n) = \begin{cases} \mathcal{P}(\text{Fib}, n) + 1 & \text{si } n = 1; \\ \mathcal{P}(\text{Fib}, n) & \text{sinon} \end{cases}$$

Néanmoins, il n'existe pas de contre-exemple connu de mot récurrent qui ne satisfasse pas l'équation (2.1). Une question naturelle semble alors être si l'on peut

étendre le théorème de Brlek et Reutenauer à l'ensemble des mots récurrents. Une telle question a déjà été posée dans un contexte légèrement différent, soit au sujet de la positivité de  $T_{\text{Id}}(\mathbf{w}, n)$  (Bucci *et al.*, 2009). Dans cet article, les auteure et auteurs disent n'avoir pu trouver ni preuve ni contre-exemple de mots récurrents pour lesquels  $T_{\text{Id}}(\mathbf{w}, n)$  est strictement négatif, alors que des contre-exemples sont connus pour les mots non-récurrents, comme celui de  $c \cdot \text{Fib}$ .

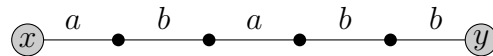


## CHAPITRE III

### LA COMPLEXITÉ PALINDROMIQUE DES ARBRES.

Le troisième chapitre délaisse l'étude des pseudo-palindromes et se concentre sur une toute autre généralisation de l'étude de la complexité palindromique : on examine ici le nombre maximal de palindromes distincts dans des arbres.

Une paire de mots  $\{u, \widehat{u}\}$  est identifiée avec un arbre composé d'une unique branche (voir figure 3.1). Ainsi, les arbres semblent être des généralisations naturelles des mots et s'interroger sur les motifs qui y apparaissent l'est tout autant.



**Figure 3.1** Un arbre filiforme représente une paire formée d'un mot et de son image miroir.

Il y a deux motivations principales : la première est intrinsèque - c'est-à-dire qu'on s'intéresse au problème en tant que finalité - alors que la seconde est plutôt liée à l'étude des plans discrets.

**Langage des arbres.** Récemment, le nombre maximal de carrés dans un arbre a été investigué (Crochemore *et al.*, 2012). Pour ce faire, les auteurs ont défini le

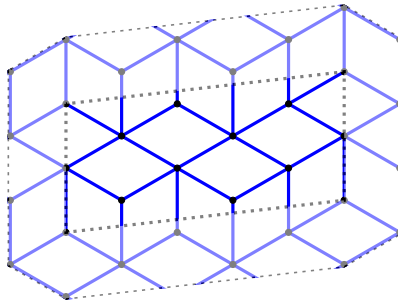
langage d'un arbre (de la même façon que dans les préliminaires de ce mémoire). Ils ont ainsi montré que le nombre maximal de carrés pour un arbre  $T$  de taille  $n$  est

$$c_1 n^{4/3} \leq \#\text{Carrés}(T) \leq c_2 n^{4/3}.$$

En comparaison, le nombre maximal de carrés dans un mot est linéaire par rapport à sa longueur (Fraenkel et Simpson, 1998). Le problème d'étudier les motifs qui apparaissent dans le langage des arbres semble intéressant et compter les palindromes est un autre problème qui permet de mieux comprendre le langage des arbres. On verra d'ailleurs que, comme pour les carrés, le nombre maximal de palindromes dans un arbre est asymptotiquement plus grand que dans un mot de même taille.

**Des motivations géométriques!** Si l'on considère les mots comme l'encodage d'objets géométriques, on peut étendre certaines définitions. Un exemple naturel est certainement les mots de Christoffel, d'abord présentés à l'exemple 1.1.2, qui codent les droites discrètes de pente rationnelle. Sébastien Labbé et Christophe Reutenauer ont d'ailleurs récemment proposé une définition des « mots » de Christoffel multi-dimensionnels (Labbé et Reutenauer, 2015). Ceux-ci sont vus comme des sous-graphes orientés d'une grille à coordonnées entières  $\mathbb{Z}^n$ . Ils représentent aussi le dessus de la *discrétisation* d'un hyperplan.

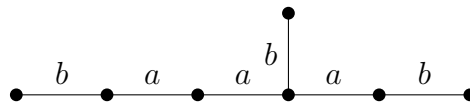
Le graphe d'adjacence de structures obtenues par symétries a été étudié par Éric Domenjoud et Laurent Vuillon dans un article où ils détaillent les graphes obtenus par clôture palindromique, de manière analogue aux mots sturmiens (de Luca, 1997) ou épisturmiens (Droubay *et al.*, 2001; Glen et Justin, 2009), aussi obtenus par clôture palindromique, et ils montrent qu'il s'agit d'arbres



**Figure 3.2** Le domaine fondamental du graphe de Christoffel associé au vecteur  $(2, 3, 5)$  (Labbé et Reutenauer, 2015).

(Domenjoud et Vuillon, 2012). Leurs palindromes ont été étudiés par les deux mêmes auteurs, conjointement avec Xavier Provençal (Domenjoud *et al.*, 2016), dans un article où ils montrent que leur nombre est celui d'arêtes plus un. Cette dernière propriété met en évidence l'analogie avec les mots sturmiens.

La figure 3.3 présente un exemple d'un arbre contenant plus de palindromes non-vides que d'arêtes, démontrant ainsi que le théorème 1.1.1 ne peut s'appliquer aux arbres. Cet exemple est d'ailleurs le plus petit qui contredise ce théorème.



**Figure 3.3** Un arbre avec 6 arêtes et 7 palindromes non-vides :  $\{a, b, aa, aaa, bab, baab, baaab\}$  (Domenjoud *et al.*, 2016).

Une particularité des arbres est, telle que l'exprime la proposition 0.0.20, que tout couple de nœuds est reliée par une unique chaîne. Ainsi, le nombre de facteurs

non-vides dans le langage d'un arbre est au plus le nombre de façons de choisir un couple de sommets distincts  $(s_i, s_j)$ , et les facteurs correspondent à l'unique chaîne de  $s_i$  à  $s_j$ . Le nombre de palindromes non-vides ne peut donc excéder le carré du nombre de sommets.

Ce chapitre exhibe d'abord une famille infinie d'arbres ayant un nombre de palindromes substantiellement plus grand que la borne indiquée par le théorème 1.1.1. On donne ensuite une valeur, à constante près, pour le nombre de palindromes dans des arbres ayant un langage d'une forme particulière. Finalement, on explique pourquoi cette valeur fonctionne en toute généralité.

**Arbres filiformes.** D'après les définitions ci-haut, on peut associer un arbre filiforme  $W$  à une paire de mots  $\{w, \tilde{w}\}$ . Supposons en effet que  $x$  et  $y$  soient ses feuilles et que les autres nœuds soient de degré 2. Alors,  $w = \pi(x, y)$  et  $\tilde{w} = \pi(y, x)$ . La taille de  $W$  est égale à  $|w| = |\tilde{w}|$ . De manière analogue,  $\text{Pal}(W) = \text{Pal}(w) = \text{Pal}(\tilde{w})$ . Le langage de  $W$  correspond à l'union des langages de  $w$  et de  $\tilde{w}$ .

*Exemple 3.0.2.* La figure 3.1 présente l'arbre filiforme associé au mot  $ababb$ . Un facteur de l'arbre est soit un facteur du mot  $\pi(x, y)$ , si les arêtes sont lues de gauche à droite, ou de  $\pi(y, x)$ , sinon.

### 3.1 Une famille d'arbres avec beaucoup de palindromes

Cette section présente une famille infinie d'arbres dont le nombre de palindromes dépasse largement sa taille. De cette façon, on établit une borne inférieure

non-triviale sur le nombre maximal de palindromes qu'un arbre peut avoir.

Pour ce faire, on présente d'abord quelques notions sur les suites de nombres, qui seront utiles pour compter les palindromes dans les arbres en « forme de peigne ».

### 3.1.1 Quelques éléments de combinatoire additive

On dit qu'une suite d'entiers est *de Sidon* si les sommes de n'importe quelle paire de ses éléments sont distinctes deux à deux (Sidon, 1932). Il existe une infinité de tels ensembles. Par exemple, les puissances de 2 (ou de n'importe quel entier au moins égal à 2) forment un ensemble de Sidon infini.

*Exemple 3.1.1.* L'ensemble  $S = \{1, 2, 4, 6, 10, 16\}$  est un ensemble de Sidon fini. En effet, les sommes de deux éléments distincts de l'ensemble sont toutes différentes :

$$\begin{array}{cccc}
 3 = 1 + 2 & 8 = 2 + 6 & 14 = 4 + 10 & 20 = 4 + 16 \\
 5 = 1 + 4 & 10 = 4 + 6 & 16 = 6 + 10 & 22 = 6 + 16 \\
 6 = 2 + 4 & 11 = 1 + 10 & 17 = 1 + 16 & 26 = 10 + 16 \\
 7 = 1 + 6 & 12 = 2 + 10 & 18 = 2 + 16 &
 \end{array}$$

**Tableau 3.1** La table d'addition de l'ensemble  $S$  montre que c'est bien un ensemble de Sidon, puisque les sommes de deux éléments distincts sont toutes différentes.

Or, bien qu'il soit relativement facile d'imaginer des ensembles satisfaisant la propriété de Sidon, en trouver qui soient denses est plus difficile.

Plusieurs propriétés des ensembles de Sidon sont connues, comme celle-ci :

**Proposition 3.1.2.** *Un ensemble est de Sidon si et seulement si les différences non-nulles des paires de ses éléments sont distinctes deux à deux.*

*Démonstration.* Soit  $S$  un ensemble de Sidon.

Par définition, pour tout  $(a, b, c, d) \in S^4$ , avec  $a \neq b$  et  $c \neq d$ ,

$$a + b = c + d \Rightarrow \{a, b\} = \{c, d\}.$$

Cette définition est équivalente à

$$a - c = d - b \Rightarrow \{a, b\} = \{c, d\}.$$

La contraposée de cet énoncé dit que si deux paires d'éléments sont distinctes, leurs différences le sont aussi, d'où l'équivalence des définitions.  $\square$

La taille maximale d'un sous-ensemble de Sidon  $S$  de  $\{1, 2, \dots, n\}$  est connue à constante près (Gowers, 2012). En effet, puisque  $S$  est de Sidon, il y a  $\frac{|S|^2}{2}$  sommes des paires d'éléments pas nécessairement distincts de  $S$ , et les sommes des termes sont toutes inférieures à  $2n$ . D'où,

$$\frac{|S|^2}{2} \leq 2n$$

et alors  $|S| \leq 2\sqrt{n}$ .

Un exemple connu d'ensemble de Sidon de cet ordre est donné par Erdős et Turán (Erdős et Turán, 1941) et consiste en la suite  $A_p$  dont les éléments sont

$$A_p = (2pk + (k^2 \pmod p))_{k=1,2,\dots,p-1}, \text{ avec } p \text{ premier.} \quad (3.1)$$

On remarque par ailleurs que, comme il existe des nombres premiers arbitrairement grands, il n'y a pas de longueur maximale pour les suites de cette famille.

**Proposition 3.1.3** ((Erdős et Turán, 1941)). *La suite  $A_p$  est de Sidon quelque soit  $p$  premier.*

*Démonstration.* Supposons qu'il existe des entiers  $i, j, k, l$  tels que  $a_i + a_j = a_k + a_l$ . Comme la somme de deux éléments de la suite est toujours la somme d'un multiple de  $2p$  et d'une valeur strictement inférieure à  $2p$ , les sommes de chacun des termes du résultat doivent être égales. Ainsi, pour avoir l'égalité des sommes, on a  $2p(i + j) = 2p(k + l)$  (ou tout simplement  $i + j = k + l$ ) et  $i^2 + j^2 \equiv k^2 + l^2 \pmod{p}$ . On peut reformuler ceci avec

$$i - k = l - j \text{ et } i^2 - k^2 \equiv l^2 - j^2 \pmod{p}.$$

Il y a alors deux cas possibles :

- Si  $i - k = l - j = 0$ , alors  $i = k$  et  $l = j$ .
- Sinon,  $i + k \equiv l + j \pmod{p}$ . En additionnant de chaque côté avec  $i - k = l - j$ , on trouve  $2i \equiv 2l \pmod{p}$  et  $2k \equiv 2j \pmod{p}$ . Or, puisque 2 ne divise pas  $p$  et que  $i, j, k$  et  $l$  sont inférieurs à  $p$ , l'égalité des paires est vérifiée.

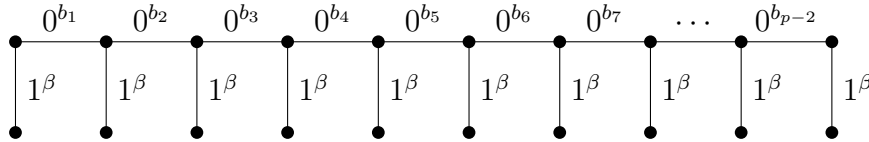
On déduit donc que  $\{i, j\} = \{k, l\}$  si les sommes des paires sont égales, et  $A_p$  est bien de Sidon. □

De plus, la suite  $A_p$  est, à constante multiplicative près, le plus dense possible. En effet, la valeur maximale d'un élément de  $A_p$  est inférieure à  $2p^2$  et  $|A_p| = p - 1$ . Donc, pour tout  $p$ ,  $A_p$  contient  $p - 1$  éléments parmi  $\{1, 2, \dots, 2p^2\}$ . Or, on sait qu'une suite de Sidon contient au plus  $2\sqrt{n}$  termes, où  $n$  est la valeur maximale de la suite. Pour  $A_p$ , la densité est  $\sqrt{8}$  fois plus petite, et ce, pour tout  $p$ . En somme, on obtient un ensemble dont la densité est d'ordre maximal.

### 3.1.2 La construction en « forme de peigne »

Notre objectif est de construire un arbre dont le langage palindromique est substantiellement plus grand que la taille de l'arbre. Dans cette section, on donne la construction d'une famille d'arbres  $\{\mathcal{Q}_{p,\beta} \mid p, \beta \in \mathbb{N}, p \text{ premier}\}$  arbitrairement grands qui contiennent au plus, à constante près,  $|\mathcal{Q}_{p,\beta}| \sqrt{|\mathcal{Q}_{p,\beta}|}$  palindromes.

On considère le « peigne »  $\mathcal{Q}_{p,\beta}$ , représenté à la figure 3.4.



**Figure 3.4** Le peigne  $\mathcal{Q}_{p,\beta}$ . Sur la figure, on note  $b_i = a_{i+1} - a_i$  où  $A$  est la suite  $A_p$  donnée à l'équation (3.1). La notation  $0^b$  sur la figure désigne une suite de  $b$  arêtes étiquetées par 0.

**Proposition 3.1.4** (Proposition 3, (Brlek *et al.*, 2015)). *La suite  $B = (b_1, \dots, b_{p-2})$  définie à la figure 3.4 est telle que la somme des termes de chaque sous-suite contiguë est distincte.*

*Démonstration.* On procède par l'absurde.

Supposons qu'il existe des indices  $k, l, m, n$  tels que

$$\sum_{i=k}^l b_i = \sum_{j=m}^n b_j.$$

Par la définition de  $B$ ,

$$\sum_{i=k}^l b_i = \sum_{i=k}^l (a_{i+1} - a_i) = a_{l+1} - a_k.$$



En procédant de la même façon, on trouve que

$$\sum_{j=m}^n b_j = a_{n+1} - a_m.$$

Il s'ensuit que

$$a_{l+1} - a_k = a_{n+1} - a_m.$$

Or, la suite  $A_p$  est de Sidon. Donc  $\{l+1, m\} = \{k, n+1\}$  et les sommes des termes des sous-suites de  $B$  sont deux à deux distinctes.  $\square$

**Lemme 3.1.5.** *Le nombre de palindromes de  $\mathcal{Q}_{p,\beta}$  est dans  $\Theta(\beta p^2)$ .*

*Démonstration.* Les palindromes non-vides de  $\mathcal{Q}_{p,\beta}$  sont de trois différentes formes. On note  $c_0$  le nombre de palindromes de la forme  $0^+$ ,  $c_1$  le nombre de ceux de la forme  $1^+$  et  $c_{101}$  le nombre de palindromes de  $1^+0^+1^+$ . On peut alors dénombrer les palindromes de  $\mathcal{Q}_{p,\beta}$  :

$$|\text{Pal}(\mathcal{Q}_{p,\beta})| = c_0 + c_1 + c_{101} + 1,$$

où l'unité est ajoutée pour le mot vide. Puis,

$$\begin{aligned} c_0 &= b_1 + b_2 + \cdots + b_{p-2} = a_{p-1} - a_1 = 2p^2 - 4p, \\ c_1 &= \beta, \\ c_{101} &= |\{\{x, y\} \mid 1^x 0^y 1^x \in \mathcal{L}(\mathcal{Q}_{p,\beta}), x, y \geq 1\}| \\ &= |\{x \mid 1 \leq x \leq \beta\}| \cdot |\{y \mid y = \sum_{i=k}^l b_i \text{ pour } 1 \leq k \leq l \leq p-2\}| \\ &= \beta(p-2) + \frac{\beta(p-2)(p-3)}{2}. \end{aligned}$$

La dernière égalité vient de la proposition 3.1.4, qui garantit que, pour chaque choix de paires d'indices, la somme est différente. Le comportement asymptotique du nombre de palindromes est déterminé par le terme dominant, soit  $\beta p^2$ .  $\square$

**Lemme 3.1.6.** *La taille de  $\mathcal{Q}_{p,\beta}$  est dans  $\Theta(p^2 + \beta p)$ .*

*Démonstration.* Le nombre d'arêtes d'étiquettes 0 est  $b_1 + b_2 + \dots + b_{p-2} = 2p^2 - 4p$ . Pour celles étiquetées par 1, il y a exactement  $p - 1$  suites d'arêtes étiquetées par 1 et leur longueur est  $\beta$ . Le nombre total d'arêtes est donc  $2p^2 - 4p + \beta(p - 1)$ .  $\square$

Pour simplifier la notation, on note  $\mathcal{Q}_p$  l'arbre  $\mathcal{Q}_{p,p}$ . On s'en sert pour exprimer le théorème suivant.

**Théorème 3.1.7** (Lemmes 4 et 5, (Brlek *et al.*, 2015)). *Le nombre maximal de palindromes de  $\mathcal{Q}_p$  est dans  $\Theta(|\mathcal{Q}_p|^{\frac{3}{2}})$ .*

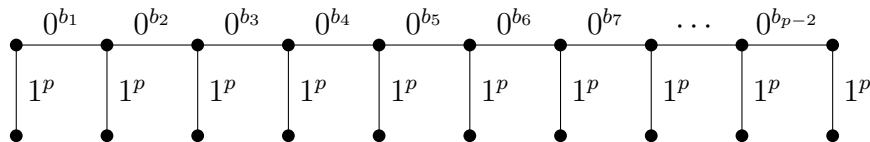
*De plus, le nombre maximal de palindromes d'un arbre  $\mathcal{Q}_{p,\beta}$  est dans  $\mathcal{O}(|\mathcal{Q}_{p,\beta}|^{\frac{3}{2}})$ .*

*Démonstration.* On se sert des lemmes 3.1.5 et 3.1.6 pour calculer  $\log_{|\mathcal{Q}_{p,\beta}|}(|\text{Pal}(\mathcal{Q}_{p,\beta})|)$ .

$$\begin{aligned} \log_{|\mathcal{Q}_{p,\beta}|}(|\text{Pal}(\mathcal{Q}_{p,\beta})|) &= \frac{\log_p(|\text{Pal}(\mathcal{Q}_{p,\beta})|)}{\log_p(|\mathcal{Q}_{p,\beta}|)} \\ &\rightarrow \frac{\log_p(\beta p^2)}{\log_p(p^2 + \beta p)} \\ &\rightarrow \frac{\log_p(\beta) + \log_p(p^2)}{\max(\log_p(p^2), \log_p(\beta p))} \\ &= \frac{\log_p(\beta) + 2}{\max(2, 1 + \log_p(\beta))}, \end{aligned}$$

qui est maximal en posant  $\beta \in \Theta(p)$ . On peut donc choisir  $\beta = p$  pour optimiser  $\log_{|\mathcal{Q}_{p,\beta}|}(|\text{Pal}(\mathcal{Q}_{p,\beta})|)$ . Alors,  $\log_{|\mathcal{Q}_p|}(|\text{Pal}(\mathcal{Q}_p)|)$  tend vers  $\frac{3}{2}$  lorsque  $p$  est grand et le nombre de palindromes de  $\mathcal{Q}_p$  est dans  $\Theta(|\mathcal{Q}_p|^{\frac{3}{2}})$ .  $\square$

*Exemple 3.1.8.* En posant  $\beta = p$ , on trouve l'arbre  $\mathcal{Q}_p$  :



**Figure 3.5** L'arbre  $\mathcal{Q}_p$ .

Ainsi, le nombre de palindromes dans l'arbre est  $\frac{p^3+p^2-4p+2}{2}$  et sa taille est  $3p^2-5p$ .  
D'où  $|\text{Pal}(\mathcal{Q}_p)| \in \Theta(|\mathcal{Q}_p|^{\frac{3}{2}})$ .

### 3.2 Vers une borne maximale

Soit  $w$  un mot. On note alors  $\Delta(w)$  son codage par blocs, c'est-à-dire la suite de longueur des blocs d'une même lettre.

*Exemple 3.2.1.* Par exemple, le mot de langue française « appelle » a pour codage par blocs  $\Delta(\text{appelle}) = 12121$ . De la même façon, pour la suite de nombres  $w = 11112111211211$ ,  $\Delta(w) = 4131212$ .

Il est facile de voir que chaque lettre de  $\Delta(w)$  représente la longueur d'un bloc, alors que la longueur de  $\Delta(w)$  peut être associée avec le nombre de blocs dans  $w$ .

On peut maintenant définir une suite infinie de familles d'arbres :

$$\mathcal{T}_k = \{\text{arbre } T \mid |\Delta(f)| \leq k \forall f \in \mathcal{L}(T)\}.$$

Pour tout entier positif  $k$ , on peut compter le nombre maximal de palindromes

d'un arbre de  $\mathcal{T}_k$  en fonction de sa taille. Pour ce faire, on définit la fonction

$$\mathcal{P}_k(n) = \max_{T \in \mathcal{T}_k, |T| \leq n} |\text{Pal}(T)|.$$

Il importe de souligner que  $\mathcal{P}_k(n)$  est une fonction croissante à la fois par rapport à  $k$  et à  $n$ . Pour le voir, on note que  $\mathcal{T}_k \subseteq \mathcal{T}_{k+1}$ .

*Exemple 3.2.2.* Les arbres  $\mathcal{Q}_{p,\beta}$ , définis à la section précédente, sont dans  $\mathcal{T}_3$ , car leur langage est un sous-ensemble de  $1^*0^*1^*$ . On peut donc dire que  $\mathcal{P}_3(n) \in \Omega(n^{\frac{3}{2}})$ . En outre, pour tout  $k > 3$ ,  $\mathcal{P}_k(n) \in \Omega(n^{\frac{3}{2}})$  par la remarque précédente.

La valeur de  $\mathcal{P}_k(n)$  est toujours au moins  $n + 1$ . Par exemple, le mot  $a^n$  contient toujours  $n + 1$  palindromes distincts (un de chaque longueur) et appartient à chacune des classes  $\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3, \dots$

Parallèlement, l'équation (0.1) indique une borne supérieure triviale pour  $\mathcal{P}_k(n)$ , puisqu'elle implique que  $\mathcal{P}_\infty(n) \in \mathcal{O}(n^2)$ .

Dans ce chapitre, on fournit la valeur asymptotique, en notation  $\Theta$ , de  $\mathcal{P}_k(n)$ , pour  $k \leq 4$ . La sous-section 3.2.3 détaille pourquoi on a conjecturé que  $\mathcal{P}_\infty(n) \in \Theta(\mathcal{P}_4(n))$ . Pour sa part, la sous-section 3.2.4 élabore une piste de résolution exploitée par les chercheurs ayant démontré la conjecture.

### 3.2.1 Arbres dans $\mathcal{T}_1$ et $\mathcal{T}_2$

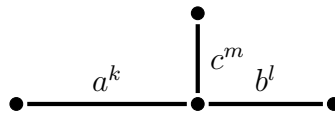
On rappelle d'abord que, par définition, chaque facteur non-vide d'un arbre de  $\mathcal{T}_2$  a soit un ou deux blocs de lettres distinctes. En d'autres termes, chaque facteur est de la forme  $a^*b^*$ , pour deux lettres  $a$  et  $b$ . Les palindromes sont donc la répétition d'une seule lettre.

Cela permet de déduire une valeur pour  $\mathcal{P}_2(n)$  :

**Proposition 3.2.3** (Proposition 2, (Brlek *et al.*, 2015)). *Le nombre maximal de palindromes pour les arbres de la famille  $\mathcal{T}_2$  est  $\mathcal{P}_2(n) = n + 1$ .*

*Démonstration.* Pour chaque lettre, le nombre de palindromes qui sont puissances de cette lettre est la longueur du plus long bloc de cette lettre. Ainsi, le nombre total de palindromes non-vides est au plus le nombre d'arêtes, et  $\mathcal{P}_2(n) = n + 1$ .  $\square$

*Exemple 3.2.4.* L'arbre de la figure 3.6 appartient à  $\mathcal{T}_2$ . Ses palindromes sont tous des répétitions de  $a$ ,  $b$  ou  $c$ . Leur nombre est au plus la taille de l'arbre, plus 1.



**Figure 3.6** Un arbre de  $\mathcal{T}_2$ .

### 3.2.2 Arbres dans $\mathcal{T}_3$ et $\mathcal{T}_4$

Les arbres de  $\mathcal{T}_3$  et  $\mathcal{T}_4$  sont les premiers exemples qui permettent de dépasser le nombre de palindromes qu'on trouve dans les mots. Par exemple, l'arbre de la figure 3.3 appartient à  $\mathcal{T}_3$ .

Dans cette partie, on montre que  $\{\mathcal{P}_3(n), \mathcal{P}_4(n)\} \subseteq \Theta(n^{\frac{3}{2}})$ . Pour se faire, on utilise la construction du peigne, qui fournit des arbres arbitrairement grands dans  $\mathcal{T}_3$  dont le nombre de palindromes est suffisant pour montrer que  $\mathcal{P}_3(n) \in \Omega(n^{\frac{3}{2}})$ . La partie plus corsée est de montrer que cette construction est optimale pour tous les arbres de  $\mathcal{T}_3$  et  $\mathcal{T}_4$ .

**La valeur de  $\mathcal{P}_4(n)$  est dans  $\Theta(n^{\frac{3}{2}})$ .**

On s'attelle alors à démontrer que la valeur asymptotique de  $\mathcal{P}_4(n)$  est atteinte par la construction du peigne.

**Théorème 3.2.5** (Théorème 7, (Brlek *et al.*, 2015)).  $\mathcal{P}_4(n) \in \Theta(n^{\frac{3}{2}})$ .

Avant de donner la preuve de ce théorème, on explique certains des arguments qui seront utilisés. On justifie d'abord pourquoi on se permet de réduire n'importe quel arbre de  $\mathcal{T}_4$  à un arbre de  $\mathcal{T}_3$  sur un alphabet binaire. On discute ensuite de certaines propriétés de ces arbres, dans le but d'établir une borne supérieure pour  $\mathcal{P}_4(n)$ .

**Lemme 3.2.6** (Lemme 8, (Brlek *et al.*, 2015)). *Pour tout arbre  $T \in \mathcal{T}_4$ , il existe un arbre  $S \in \mathcal{T}_3$  sur un alphabet binaire et satisfaisant les deux conditions suivantes :*

- $|S| \leq |T|$ ;
- $\frac{|\text{Pal}(T)|}{|\Sigma|^2} - |T| \leq |\text{Pal}(S)| \leq |\text{Pal}(T)|$ .

*Démonstration.* Il y a deux possibilités pour  $T$ . Soit il ne contient aucun facteur avec trois blocs de lettres distinctes qui commence et termine par la même lettre, soit il en contient. Dans le premier cas, les palindromes sont alors tous des répétitions d'une seule lettre. Supposons que  $a$  soit la lettre qui compose le plus long palindrome. Elle n'est possiblement pas unique (ce n'est pas important). Soit  $S$  l'arbre composé uniquement de la plus longue chaîne qui soit étiquetée seulement par la lettre  $a$ . Alors,  $|\text{Pal}(T)| \leq |\Sigma| |\text{Pal}(S)| \leq |\Sigma| |\text{Pal}(T)|$ .

Dans l'autre cas, on peut supposer que  $a$  et  $b$  sont les lettres telles que le couple

$(a, b)$  maximise  $|\mathcal{L}(T) \cap \text{Pal}(a^+b^+a^+)|$ . On définit l'ensemble des arêtes contribuant aux palindromes de  $a^+b^+a^+$

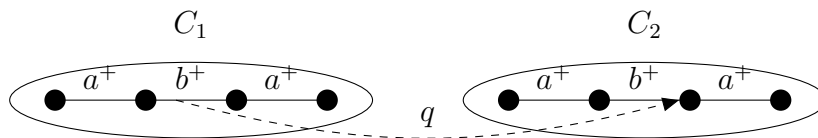
$$E_S = \cup( p(u, v) \mid \pi(u, v) \in \text{Pal}(a^+b^+a^+) )$$

et soit  $S$  le sous-graphe de  $T$  qui contient exactement les arêtes des chaînes de  $E_S$  et les nœuds qui y sont adjacents. Il y a alors trois choses à prouver :

- $S$  est un arbre : Puisque  $S$  est un sous-graphe de  $T$ , il ne peut y avoir de cycle. Il faut cependant montrer qu'il est connexe. Supposons que  $S$  a deux composantes connexes, appelées  $C_1$  et  $C_2$ . Comme elles sont dans  $S$ ,  $\mathcal{L}(C_1) \subseteq a^*b^*a^*$  et  $C_1$  a au moins un facteur de  $a^+b^+a^+$ . Il en est de même pour  $C_2$ .

Comme  $T$  est un arbre, il existe une chaîne de  $T \setminus S$  qui rejoint  $C_1$  à partir de  $C_2$ . On nomme cette chaîne  $q$ .

Il y a des chaînes dans  $C_1$  et dans  $C_2$  qui partent d'une extrémité de  $q$  et dont la trace est  $a^*b^+a^+$ . Ainsi, si  $w$  est la trace de  $q$ ,  $T$  contient au moins un facteur  $f \in a^+b^+a^*wa^*b^+a^+$ . Par hypothèse,  $T \in \mathcal{T}_4$  et donc tout facteur de  $T$  contient au plus quatre blocs de lettres distinctes. Alors  $f \in a^+b^+wb^+a^+$ , avec  $w \in b^*$ , ce qui contredit le fait que  $q$  appartient à  $T \setminus S$ .



**Figure 3.7** La connexité de l'arbre  $S$  est garantie par le fait que c'est une restriction d'un arbre de  $\mathcal{T}_4$ .

- $S$  est un arbre de  $\mathcal{T}_3$  étiqueté par un alphabet binaire : par construction,  $S$  ne contient que des arêtes étiquetées par  $a$  ou  $b$ . Il n'a de plus aucune

feuille adjacente à une arête  $b$ . Cela a pour conséquence que si  $S$  contient un facteur  $f \in a^+b^+a^+b^+$ ,  $f$  peut être étendu à  $f' \in a^+b^+a^+b^+a^+$ , qui ne peut apparaître dans un arbre de  $\mathcal{T}_4$ .

- $|\text{Pal}(S)| \geq \frac{1}{|\Sigma|^2}|\text{Pal}(T)| - |T|$  : on a choisi le couple  $(a, b)$  de manière à maximiser le nombre de palindromes qui contiennent deux lettres distinctes dans  $T$ . Le nombre de palindromes qui sont répétitions d'une lettre est au plus  $|T|$ . Ainsi,

$$\frac{1}{|\Sigma|^2}|\text{Pal}(T)| - |T| \leq |\text{Pal}(S)| \leq |\text{Pal}(T)|. \quad \square$$

**Lemme 3.2.7** (Lemme 9, (Brlek et al., 2015)). *Pour tout  $T \in \mathcal{T}_3$ ,  $T$  ne peut contenir à la fois des facteurs dans  $a^+b^+a^+$  et dans  $b^+a^+b^+$ .*

*Démonstration.* On procède par contradiction. Supposons qu'il existe dans  $T$  quatre nœuds  $u, v, x, y$  tels que  $\pi(u, v) \in a^+b^+a^+$  et  $\pi(x, y) \in b^+a^+b^+$ . Puisque  $T$  est un arbre, il existe une unique chaîne entre deux nœuds. En particulier, il y a une chaîne de  $w \in \{u, v\}$  à  $w' \in \{x, y\}$  qui contient un facteur de  $a^+b^+a^+\Sigma^*b^+$ , ce qui contredit l'hypothèse que  $T \in \mathcal{T}_3$ .  $\square$

On définit maintenant la restriction d'un arbre  $T$ , notée  $\mathcal{R}_a(T)$ , aux seules arêtes de  $T$  associées à la lettre  $a$  et aux nœuds qui leur sont adjacents.

**Lemme 3.2.8** (Lemme 10, (Brlek et al., 2015)). *Soit  $T \in \mathcal{T}_3$ . Il existe au moins une lettre  $a \in \Sigma$  telle que  $\mathcal{R}_a(T)$  est connexe.*

*Démonstration.* Si  $T$  ne contient aucun facteur ayant au moins deux lettres distinctes et qui commence et termine par la même lettre, c'est-à-dire de la forme



$b^+a^+b^+$ , alors  $\mathcal{R}_a(T)$  est connexe pour toute lettre  $a$ .

Sinon, supposons que le facteur  $f \in b^+a^+b^+$  est dans  $T$ . S'il existe une composante connexe de  $\mathcal{R}_a(T)$  disjointe de  $f$ , celle-ci doit être reliée par une chaîne  $p$  dans  $T$  qui ne contienne pas que des  $a$ . Soit  $w$  la trace de  $p$ . Alors  $awb^*a^+b^+ \in \mathcal{L}(T)$ , ce qui contredit le fait que  $T \in \mathcal{T}_3$ .  $\square$

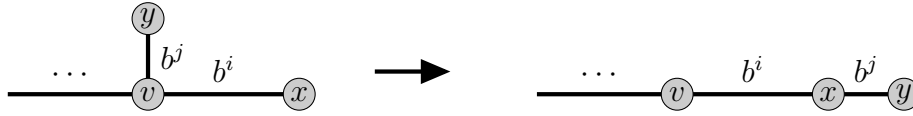
Étant donné un nœud  $s$  dans un arbre, on dit que  $s$  est une *rupture sur la lettre*  $a$  si le degré de  $s$  est au moins 3 et qu'il y a au moins deux arêtes adjacentes à  $s$  qui sont étiquetées par  $a$  (voir figure 3.8).

**Lemme 3.2.9** (Lemme 11, (Brek et al., 2015)). *Soit  $T$  un arbre de  $\mathcal{T}_3$ . Alors il existe un arbre  $T'$  de taille  $|T|$  tel que  $\mathcal{L}(T) \subseteq \mathcal{L}(T')$  et il existe une lettre  $a \in \Sigma$  telle que toutes les ruptures de  $T'$  sont sur la lettre  $a$ .*

*Démonstration.* Si  $T$  appartient à  $\mathcal{T}_2$ , on applique la transformation suivante à chacune des branches. Sinon, on suppose qu'un facteur de  $b^+a^+c^+$  apparaît dans  $T$ . Il est ici possible que  $b = c$ . On permettra alors les ruptures sur la lettre  $a$ .

Soit  $v$  un nœud de  $T$  qui soit une rupture sur une lettre  $b \neq a$ . S'il n'en existe pas, on pose  $T' = T$  et on a terminé. Par hypothèse sur  $T$ , il existe, à partir de  $v$ , au moins deux chaînes étiquetées uniquement par des  $b$  et menant à des feuilles, qui sont notées respectivement  $x$  et  $y$ .

Sans perte de généralité, on peut supposer que  $|\pi(v, x)| \geq |\pi(v, y)|$ . Alors, les mots qui ont un préfixe de  $\pi(v, y)$  comme suffixe sont un sous-ensemble de ceux ayant comme suffixe un préfixe de  $\pi(v, x)$ . Par conséquent, les seules



**Figure 3.8** La destruction d'une rupture sur la lettre  $b$ .

contributions au langage de  $T$  qui nécessitent  $\pi(v, y)$  sont les chaînes qui passent à la fois par des arêtes de  $\pi(v, x)$  et de  $\pi(v, y)$ . Les mots qui nécessitent les deux branches ne sont composés que de  $b$  et sont de longueur au plus  $|\pi(v, x)| + |\pi(v, y)|$ .

On peut déplacer les arêtes entre  $v$  et  $y$  à l'autre extrémité de  $x$ ; on construit alors un arbre dont le langage contient celui de  $T$  et qui a le même nombre de nœuds. Finalement, on applique cette procédure jusqu'à ce qu'il ne reste des ruptures que sur la lettre  $a$ . L'arbre ainsi créé est  $T'$ .  $\square$

On peut maintenant démontrer le théorème principal.

*Démonstration du théorème 3.2.5 :  $\mathcal{P}_4(n) \in \Theta(n^{\frac{3}{2}})$ .*

Soit  $T$  un arbre de  $\mathcal{T}_4$ .

Par hypothèse, chaque facteur de  $T$  contient au plus quatre blocs de lettres.

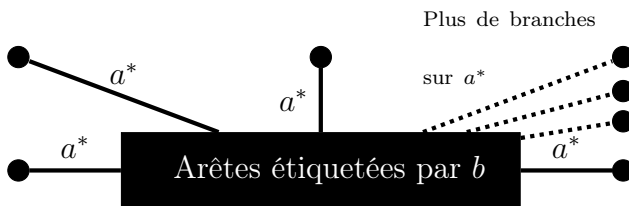
1. Soit  $S \in \mathcal{T}_3$  un arbre tel que  $|S| \leq |T|$ ,  $\mathcal{L}(S) \subseteq \{a, b\}^*$  et  $\frac{|\text{Pal}(T)|}{|\Sigma|^2} - |T| \leq |\text{Pal}(S)| \leq |\text{Pal}(T)|$ . Par le lemme 3.2.6, on sait qu'un tel arbre existe.

On sait de plus, par le lemme 3.2.7, que  $S$  peut contenir des facteurs de  $a^+b^+a^+$ , mais pas de  $b^+a^+b^+$ .

2. Par le lemme 3.2.9, il existe un arbre  $S'$  tel que  $|S'| = |S|$ ,  $\mathcal{L}(S) \subseteq \mathcal{L}(S')$  et sans rupture sur la lettre  $a$ .

3. Finalement, on compte les palindromes dans  $S'$ . Ceux qui sont non-vides sont soit dans  $b^+$ ,  $a^+$  ou  $a^+b^+a^+$ . Pour les palindromes qui sont des répétitions d'une seule lettre, leur nombre est borné par  $|S'| + 1$ . On se concentre donc sur les palindromes de  $a^+b^+a^+$ . On note ce nombre  $c_{aba}$ . On montre que  $c_{aba} \leq 2n\sqrt{n}$  lorsque  $n = |S'|$ .

Étant donné que  $S'$  n'admet pas de rupture sur la lettre  $a$ , chaque composante connexe de  $R_a(S')$  est une branche filiforme entre une feuille et  $R_b(S')$ . On nomme ces composantes connexes  $u_1, \dots, u_m$ . De plus, par le lemme 3.2.8, on sait que  $R_b(S')$  est connexe.



**Figure 3.9** La forme de l'arbre  $S'$ .

Soient  $u_i$  et  $u_j$  deux branches de  $S'$ . De façon à généraliser la notation pour une chaîne entre deux nœuds, on note  $\pi(u_i, u_j)$  le mot défini par l'unique chaîne de  $u_i$  à  $u_j$  dans  $R_b(S')$ . Soit  $l$  l'entier tel que  $\pi(u_i, u_j) = b^l$  et supposons que  $|u_i| \leq |u_j|$ . Alors, pour tout nœud  $v$  dans  $u_i$ , il existe un unique nœud  $w$  de  $u_j$ , tel que le mot  $\pi(v, w) = a^k b^l a^k$  est un palindrome.

De plus, si  $|u_i| < |u_j|$ , il y a des nœuds de  $u_j$  qui ne peuvent être couplés à un

autre nœud de  $u_i$  pour former un palindrome. À partir de cette observation, on trouve une première borne supérieure :

$$c_{aba} \leq \sum_{1 \leq i < j \leq m} \min(|u_i|, |u_j|). \quad (3.2)$$

Une autre façon de borner  $c_{aba}$  est de compter les palindromes de  $a^+b^+a^+$  en fonction de la longueur du bloc de  $b$ . Pour chaque longueur  $l$  de 1 à  $n$ , il peut exister une paire de branches  $\{u_i, u_j\}$  qui produise le plus de palindromes ayant pour facteur central  $b^l$ . Ceci permet d'obtenir une deuxième borne supérieure :

$$c_{aba} \leq \sum_{l=1}^n \max_{\substack{1 \leq i < j \leq m \\ \pi(u_i, u_j) = b^l}} (\min(|u_i|, |u_j|)). \quad (3.3)$$

Pour finalement obtenir la borne désirée sur  $c_{aba}$ , on combine les deux dernières bornes obtenues. Soit  $U'$  l'ensemble des « longues » branches étiquetées par  $a$  :  $U' = \{i \mid |u_i| \geq \sqrt{n}\}$ . Comme  $n$  est la taille de  $S'$ ,  $|U'| \leq \sqrt{n}$  et la longueur moyenne de ces branches est bornée par  $n/|U'|$ . En appliquant la borne à l'équation (3.2) aux palindromes formés de deux longues branches de  $a$ , on obtient un plafond pour le nombre de ces derniers palindromes :

$$\sum_{1 \leq i < j \leq m, \{i, j\} \subseteq U'} \min(|u_i|, |u_j|) \leq \frac{|U'|(|U'| - 1)}{2} \frac{n}{|U'|} \leq n\sqrt{n}. \quad (3.4)$$

On remarque qu'une partie de l'équation vient du fait que la somme des minimums de chaque paire d'un ensemble de nombres positifs est toujours inférieure ou égale au nombre de paires multipliées par la moyenne.

Finalement, il reste à compter le nombre de palindromes qui ont de « courtes » suites de  $a$ , c'est-à-dire ceux qui sont définis par des paires de branches  $\{u_i, u_j\}$

dont soit  $i$  ou  $j$  n'est pas dans  $U'$ . Dans ce cas, on trouve toujours que  $\min(|u_i|, |u_j|) < \sqrt{n}$ . Le nombre de palindromes est alors :

$$\sum_{l=1}^n \max_{\substack{1 \leq i < j \leq m \\ \pi(u_i, u_j) = b^l, \{i, j\} \notin U'}} (\min(|u_i|, |u_j|)) < n\sqrt{n}. \quad (3.5)$$

Comme chaque palindrome de  $S'$  est compté par les équations (3.4) ou (3.5), on obtient, par sommation, que  $c_{aba} < 2n\sqrt{n} = 2|S'|^{\frac{3}{2}}$ .

On peut donc en déduire que, pour tout arbre  $T$  dans  $\mathcal{T}_4$ , le nombre de palindromes est borné par

$$|\text{Pal}(T)| \leq |\Sigma|^2(|\text{Pal}(S)| + |T|) < |\Sigma|^2(2|S'|^{\frac{3}{2}} + |T|) \leq |\Sigma|^2(2|T|^{\frac{3}{2}} + |T|).$$

Comme l'alphabet est de taille fixée, c'est suffisant pour prouver  $\mathcal{P}_4(n) \in \mathcal{O}(n^{\frac{3}{2}})$ . En combinant ce résultat avec celui donné plus haut (théorème 3.1.7), on obtient que  $\mathcal{P}_3(n)$  et  $\mathcal{P}_4(n)$  sont dans  $\Theta(n^{\frac{3}{2}})$ .  $\square$

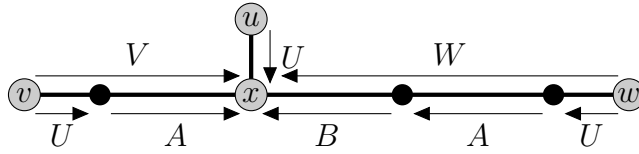
### 3.2.3 Hypothèses pour la construction d'arbres avec un grand nombre de palindromes distincts

Soit  $T$  un arbre qui maximise le nombre de palindromes pour sa taille. On peut s'attendre à ce qu'il contienne beaucoup de triplets de nœuds  $(u, v, w)$  tels que  $\pi(u, v)$ ,  $\pi(u, w)$  et  $\pi(v, w)$  sont tous des palindromes. Supposons que c'est le cas et on définit  $T'$  comme la restriction de  $T$  aux chaînes entre  $u$ ,  $v$  et  $w$ .

Il y a deux possibilités. Soit  $T'$  est un arbre filiforme, soit  $T'$  a trois feuilles ( $u$ ,  $v$  et  $w$ ) et un unique nœud de degré 3. Le premier cas est sans intérêt particulier,

puisqu'il est équivalent à celui des mots. Le second implique une structure particulière sur les facteurs  $\pi(u, v)$ ,  $\pi(u, w)$  et  $\pi(v, w)$ . On se concentre sur ce cas et on note  $x$  l'unique nœud de degré 3 dans  $T'$ .

Notons  $U = \pi(u, x)$ ,  $V = \pi(v, x)$ ,  $W = \pi(w, x)$  et, sans perte de généralité, on suppose que  $|U| \leq |V| \leq |W|$ . Comme le montre la figure 3.10,  $U\tilde{V}$ ,  $U\tilde{W}$  et  $V\tilde{W}$  sont des palindromes.



**Figure 3.10** La structure de l'arbre  $T'$ . Le fait que  $U\tilde{V}$ ,  $U\tilde{W}$  et  $V\tilde{W}$  soient des palindromes force que  $V$  débute par  $U$  alors que  $W$  débute à la fois par  $U$  et  $V$ .

Soit  $A$  le suffixe de  $V$  de longueur  $|V| - |U|$ . Par hypothèse,  $U\tilde{V}$  est un palindrome, et alors  $V = UA$ , ce qui veut dire que  $A$  en est aussi un. De la même manière, soit  $B$  le suffixe de longueur  $|W| - |V|$  de  $W$ . Cela implique que  $W = VB = UAB$ , et à la fois  $B$  et  $AB$  sont des palindromes.

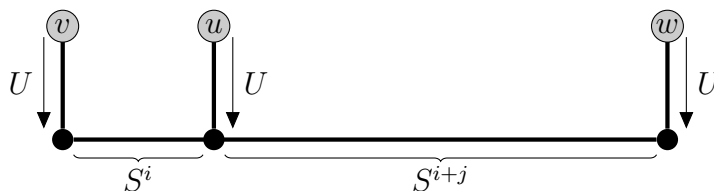
La prochaine proposition affirme que  $ABA$  est périodique et que sa période est au plus le pgcd de la différence des longueurs des trois chaînes entre  $u$ ,  $v$  et  $w$ . Plus formellement, on définit

$$p = \text{pgcd} \left( \underbrace{|\pi(u, w)| - |\pi(u, v)|}_{=|B|}, \underbrace{|\pi(v, w)| - |\pi(u, v)|}_{=|A|+|B|}, \underbrace{|\pi(v, w)| - |\pi(u, w)|}_{=|A|} \right).$$

**Proposition 3.2.10** (Proposition 13, (Brllek *et al.*, 2015)). *Il existe un mot  $S$  et deux entiers  $i, j$  tels que  $|S|$  divise  $p$ , et  $A = S^i$  et  $B = S^j$ .*

*Démonstration.* Comme  $A$ ,  $B$  et  $AB$  sont palindromes,  $AB = \widetilde{AB} = \widetilde{B}\widetilde{A} = BA$ . Ainsi, par la proposition 0.0.15, il existe un mot  $S$  tel que  $A = S^i$  et  $B = S^j$  et  $|S|$  divise  $\text{pgcd}(|A|, |B|)$ . Par construction,  $\text{pgcd}(|A|, |B|) = p$ .  $\square$

Par la proposition ci-dessus, on déduit qu'un triplet de nœuds non-alignés tels que la chaîne d'un nœud à un autre est un palindrome force une structure locale isomorphe à celle du peigne de la sous-section 3.1.2. C'est ce qu'illustre la figure 3.11.



**Figure 3.11** Un triplet de nœuds dont les chaînes entre chaque paire forme un palindrome est isomorphe à une partie du peigne.

De manière plus générale, supposons qu'un arbre contienne  $m$  feuilles  $(u_i)_{1 \leq i \leq m}$ , et que chaque  $\pi(u_i, u_j)$  est un palindrome. Soit  $T'$  la restriction de cet arbre aux chaînes qui joignent ces feuilles et, pour chaque  $i$ , soit  $v_i$  le premier nœud de degré strictement supérieur à 2 accessible à partir de la feuille  $u_i$  dans  $T'$ . En appliquant la proposition ci-dessus à chaque triplet  $(u_i, u_j, u_k)$ , pour tout  $i \neq j$ , le mot  $\pi(u_i, u_j)$  est de la forme  $\pi(u_i, u_j) = US^i\widetilde{U}$ , où  $|U| = \min_i(\pi(u_i, v_i))$  et  $|S|$  divise  $\text{pgcd}_{i \neq j, k \neq l} (|\pi(u_i, u_j)| - |\pi(u_k, u_l)|)$ .

Qui plus est, de façon à maximiser le nombre de palindromes par rapport à la taille de l'arbre, il semble avisé de choisir pour  $S$  une lettre. C'est évidemment possible puisque la seule condition pour la longueur de  $S$  est de diviser la différence de toutes les longueurs entre les chaînes palindromes d'une feuille à une autre.

On obtient de la sorte un arbre analogue à ceux présentés à la sous-section 3.1.2,

$\mathcal{Q}_{p,\beta}$ . Pour ces arbres,  $|\text{Pal}(\mathcal{Q}_p)| \in \Theta(|\mathcal{Q}_p|^{\frac{3}{2}})$  et  $\mathcal{Q}_p$  est l'arbre qui maximise cette valeur parmi  $\{\mathcal{Q}_{p,\beta}\}$ . D'où on obtient la conjecture suivante :

**Conjecture 3.2.11.** (*Brlek et al., 2015*)  $\mathcal{P}_\infty(n) \in \Theta(n^{\frac{3}{2}})$ .

### 3.2.4 La résolution du problème général

Après la publication de la conjecture 3.2.11, des chercheurs ont démontré celle-ci (Gawrychowski *et al.*, 2015). Pour ce faire, ils ont utilisé une technique similaire à celle ayant permis de résoudre le problème du nombre maximal de carrés dans un arbre (Crochemore *et al.*, 2012).

Ainsi, les auteurs de ces articles identifient le langage de n'importe quel arbre à celui d'un arbre de taille au plus deux fois supérieure. Ils associent à chaque arbre un arbre *orienté*, au sens d'un graphe orienté sans cycle, dont tous les chemins plus grands qu'une certaine longueur passent par un nœud distingué. Avec une approche récursive, ils montrent que si le nombre de carrés de n'importe quel arbre *double et déterministe* de taille  $n$  est  $n^{\frac{4}{3}}$ , alors le nombre de carrés de n'importe quel arbre de taille  $n$  est au plus  $n^{\frac{4}{3}}$ . Ils font de même pour démontrer que le nombre de palindromes dans n'importe quel arbre de taille  $n$  est au plus  $n^{\frac{3}{2}}$  lorsqu'il en est ainsi pour les arbres doubles et déterministes.

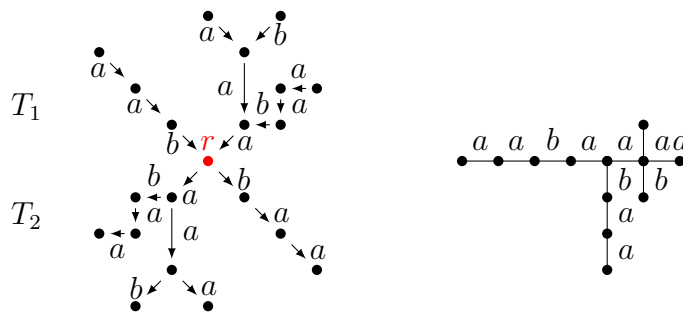
### Arbres doubles et déterministes

Tels qu'introduits dans (Crochemore *et al.*, 2012), un *arbre double*  $D = (T_1, T_2, r)$  est un arbre étiqueté composé de deux arbres disjoints  $T_1$  et  $T_2$ , sauf pour un nœud en commun,  $r$ . La taille de  $D$  est donc  $|D| = |T_1| + |T_2| - 1$ . On définit les chemins de  $D$  comme seulement les chemins simples débutant dans  $T_1$  et terminant dans  $T_2$ .



On peut donc associer à chaque arbre non-vide  $T$  un arbre double  $D = (T, T, r)$ , où  $r$  est un nœud quelconque de l'arbre  $T$ . Cet arbre est de taille  $2|T| - 1$ .

On dit qu'un tel arbre est *déterministe* si les arcs (arêtes orientées) sortant d'un nœud de  $T_1$  (différent de  $r$ ) sont étiquetés par des lettres distinctes. La figure 3.12 illustre un tel arbre.



**Figure 3.12** Un arbre double et déterministe (à gauche) dont le langage est relié à celui de l'arbre de droite.

Le résultat suivant suggère l'intérêt de déterminer les arbres doubles.

**Lemme 3.2.12** (Lemme 3, (Crochemore *et al.*, 2012)). *Il est possible de déterminer n'importe quel arbre double sans altérer son langage (au sens du langage d'un arbre double, c'est-à-dire l'ensemble des mots allant de  $T_1$  vers  $T_2$ ) et sans augmenter le nombre de nœuds.*

À partir de ce lemme, et avec un peu de travail, on peut déduire la proposition suivante :

**Proposition 3.2.13** (Preuve du théorème 8, (Gawrychowski *et al.*, 2015); analogue au lemme 4, (Crochemore *et al.*, 2012)). *Supposons qu'il existe une constante  $c$  telle que le nombre de palindromes formés par des chemins passant par n'importe quel sommet distingué dans n'importe quel arbre  $D = (T_1, T_2, r)$  double et déterministe de taille  $n$  est inférieure à  $cn\sqrt{n}$ . Alors,  $\mathcal{P}_\infty(n) \in \Theta(n\sqrt{n})$ .*

La preuve est analogue à celle du lemme 4 donnée dans (Crochemore *et al.*, 2012), qui la présente pour le nombre de carrés dans les arbres.

L'apport significatif de l'article est cependant de montrer que l'hypothèse de la précédente proposition est satisfaite. Afin d'obtenir ce résultat, ses auteurs comparent les arbres doubles et déterministes aux arbres « en forme de colonne vertébrale », qui contiennent un chemin distingué satisfaisant certaines propriétés quant à la périodicité de sa trace. Cela leur permet de conclure avec ce lemme et le théorème exprimé plus haut comme conjecture (conjecture 3.2.11).

**Lemme 3.2.14** (Lemme 7, (Gawrychowski *et al.*, 2015)). *Tout arbre double et déterministe contient au plus  $cn\sqrt{n}$  palindromes, pour une certaine constante  $c$ .*

**Théorème 3.2.15** (Théorème 8, (Gawrychowski *et al.*, 2015)).  $\mathcal{P}_\infty(n) \in \Theta(n^{\frac{3}{2}})$ .

*Démonstration.* Le lemme 3.2.14 démontre que l'hypothèse de la proposition 3.2.13 est satisfaite. Ainsi,  $\mathcal{P}_\infty(n) \in \Theta(n^{\frac{3}{2}})$ . □

## CONCLUSION

Ce mémoire se voulait un survol de certains résultats sur les palindromes et de leurs généralisations. C'est pourquoi les chapitres 1 et 2 ont mis l'accent sur les  $\sigma$ -palindromes. On a alors défini et calculé le  $\sigma$ -défaut, puis démontré qu'il interagissait avec les complexités palindromique et en facteurs. Au chapitre 3, c'était aussi le décompte des palindromes qui était à l'honneur, mais cette fois dans les arbres. Cette structure englobe celle d'un mot, qui est vu comme un arbre filiforme (à la différence près de l'orientation).

### Problèmes ouverts

Certains problèmes sont cependant toujours ouverts :

- On a largement étudié le nombre maximal de  $\sigma$ -palindromes qu'un mot, ou un arbre, peut contenir. Récemment, Gabriele Fici et Luca Zamboni ont investigué le nombre minimal de palindromes dans un mot (Fici et Zamboni, 2013). Ils ont obtenu un résultat général, ainsi qu'un résultat propre aux mots apériodiques, fermés par miroir ou sur un alphabet binaire. Un travail similaire pourrait sûrement être fait pour les  $\sigma$ -palindromes.
- Qu'arrive-t-il si on omet l'hypothèse que  $\sigma$  est une permutation involutive? La fonction  $\theta$  n'est plus involutive. Par contre, si  $\sigma$  est une permutation quelconque des lettres,  $\theta$  est d'ordre fini. On pourrait alors étudier les

propriétés combinatoires des pseudopalindromes ainsi générés.

- À la fin du chapitre 2, il est mentionné qu'il est toujours inconnu si l'équation (2.1) est valide pour n'importe quel mot récurrent. L'absence de structure propre à l'ensemble des mots récurrents semble toutefois rendre ce problème difficile à résoudre.
- Dans le but d'étudier les arbres, il pourrait être intéressant de catégoriser les arbres en fonction de familles, comme dans le cas des mots. Pour ce faire, on pourrait par exemple utiliser ce qui est connu sur la complexité en facteurs des mots. Ainsi, un *arbre ultimement périodique* serait déterminé par une complexité bornée par une constante. En revanche, on ne pourrait ainsi que classifier les arbres infinis, qui n'ont pas fait l'objet des travaux de ce mémoire.
- L'étude des palindromes parmi les chaînes simples étiquetées d'un graphe quelconque pourrait être envisageable. Cependant, ce problème se révélerait probablement plus complexe que celui des arbres.

## BIBLIOGRAPHIE

- Adamczewski, B. et Bugeaud, Y. (2007). Palindromic continued fractions. *Université de Grenoble. Annales de l'Institut Fourier*, 57(5), 1557–1574.
- Adamczewski, B. et Bugeaud, Y. (2010). Transcendence and Diophantine approximation. In *Combinatorics, automata and number theory*, volume 135 de *Encyclopedia of Mathematics and its Applications*. 410–451. Cambridge University Press, Cambridge.
- Adamczewski, B., Bugeaud, Y. et Luca, F. (2004). Sur la complexité des nombres algébriques. *Comptes Rendus Mathématique. Académie des Sciences. Paris*, 339(1), 11–14.
- Allouche, J.-P. (1997). Schrödinger operators with Rudin-Shapiro potentials are not palindromic. *Journal of Mathematical Physics*, 38, 1843–1848.
- Allouche, J.-P., Baake, M., Cassaigne, J. et Damanik, D. (2003). Palindrome complexity. *Theoretical Computer Science*, 292(1), 9–31.
- Allouche, J.-P. et Shallit, J. (1999). The ubiquitous Prouhet-Thue-Morse sequence. In *Sequences and their applications* 1–16.
- Allouche, J.-P. et Shallit, J. (2000). Sums of digits, overlaps, and palindromes. *Discrete Mathematics & Theoretical Computer Science*, 4(1), 1–10.
- Allouche, J.-P. et Shallit, J. (2003). *Automatic sequences*. Cambridge University Press.
- Anne, V., Zamboni, L. et Zorca, I. (2005). Palindromes and pseudo-palindromes in episturmian and pseudo-palindromic infinite words. Dans S. Brlek et C. Reutenauer (dir.). *Proceedings of Words, 5th International Conference On Words*, volume 36, 91–100. Publications du LaCIM.
- Baláži, P., Masáková, Z. et Pelantová, E. (2007). Factor versus palindromic complexity of uniformly recurrent infinite words. *Theoretical Computer Science*,

380(3), 266–275.

Balková, L., Pelantová, E. et Starosta, Š. (2011a). Corrigendum : On Brlek-Reutenauer conjecture. *Theoretical Computer Science*, 412(41), 5649–5655.

Balková, L., Pelantová, E. et Starosta, Š. (2011b). On Brlek-Reutenauer conjecture. *Theoretical Computer Science*, 412(41), 5649–5655.

Balková, L., Pelantová, E. et Starosta, Š. (2013). Proof of the Brlek-Reutenauer conjecture. *Theoretical Computer Science*, 475, 120–125.

Bašić, B. (2013). On highly potential words. *European Journal of Combinatorics*, 34, 1028–1039.

Berstel, J., Lauve, A., Reutenauer, C. et Saliola, F. V. (2009). *Combinatorics on words*, volume 27 de *CRM Monograph Series*. American Mathematical Society.

Berstel, J. et Perrin, D. (2007). The origins of combinatorics on words. *European Journal of Combinatorics*, 28(3), 996–1022.

Blondin Massé, A. (2008). *Sur le défaut palindromique des mots infinis*. Université du Québec à Montréal, mémoire de maîtrise.

Blondin Massé, A., Brlek, S., Frosini, A., Labbé, S. et Rinaldi, S. (2008a). Reconstructing words from a fixed palindromic length sequence. Dans G. Ausiello, J. Karhumäki, G. Mauri, et C.-H. L. Ong (dir.). *5th IFIP International Conference On Theoretical Computer Science - TCS 2008*, volume 273 de *IFIP*, 101–114.

Blondin Massé, A., Brlek, S., Garon, A. et Labbé, S. (2008b). Combinatorial properties of  $f$ -palindromes in the Thue-Morse sequence. *Pure Mathematics and Applications*, 19(2-3), 39–52.

Blondin Massé, A., Brlek, S. et Labbé, S. (2008c). Palindromic lacunas of the Thue-Morse word. Dans *Proceedings of GASCom 2008*, 53–67.

Brlek, S. (1989). Enumeration of factors in the Thue-Morse word. *Discrete Applied Mathematics*, 24(1-3), 83–96.

Brlek, S., Hamel, S., Nivat, M. et Reutenauer, C. (2004). On the palindromic complexity of infinite words. *International Journal on Foundation of Computer Science*, 15(2), 293–306.

Brlek, S. et Lafrenière, N. (2014). Reconstructing words from a  $\sigma$ -palindromic language. *Fundamenta Informaticae*, 135(1-2), 59–72.

Brlek, S., Lafrenière, N. et Provençal, X. (2015). Palindromic complexity of trees.

- Dans *Developments in Language Theory - 19th International Conference, DLT 2015*, 155–166.
- Brek, S. et Reutenauer, C. (2011). Complexity and palindromic defect of infinite words. *Theoretical Computer Science*, 412(4-5), 493–497.
- Bucci, M., De Luca, A., Glen, A. et Zamboni, L. Q. (2009). A connection between palindromic and factor complexity using return words. *Advances in Applied Mathematics*, 42(1), 60–74.
- Crochemore, M., Iliopoulos, C. S., Kociumaka, T., Kubica, M., Radoszewski, J., Rytter, W., Tyczynski, W. et Walen, T. (2012). The maximum number of squares in a tree. Dans *Combinatorial Pattern Matching - 23rd Annual Symposium, CPM 2012*, 27–40.
- Damanik, D. et Zare, D. (2000). Palindrome complexity bounds for primitive substitution sequences. *Discrete Mathematics*, 222(1-3), 259–267.
- de Luca, A. (1997). Sturmian words: Structure, combinatorics, and their arithmetics. *Theoretical Computer Science*, 183(1), 45–82.
- de Luca, A. et de Luca, A. (2006). Pseudopalindrome closure operators in free monoids. *Theoretical Computer Science*, 362(1-3), 282–300.
- Domenjoud, E., Provençal, X. et Vuillon, L. (2016). Palindromic language of thin discrete planes. *Theoretical Computer Science*.
- Domenjoud, E. et Vuillon, L. (2012). Geometric palindromic closure. *Uniform Distribution Theory*, 7(2), 109–140.
- Droubay, X., Justin, J. et Pirillo, G. (2001). Episturmian words and some constructions of de Luca and Rauzy. *Theoretical Computer Science*, 255(1-2), 539–553.
- Droubay, X. et Pirillo, G. (1999). Palindromes and Sturmian words. *Theoretical Computer Science*, 223(1-2), 73–85.
- Erdős, P. et Turán, P. (1941). On a problem of Sidon in additive number theory, and on some related problems. *Journal of the London Mathematical Society. Second Series*, 16, 212–215.
- Fici, G. et Zamboni, L. Q. (2013). On the least number of palindromes contained in an infinite word. *Theoretical Computer Science*, 481, 1–8.
- Fine, N. J. et Wilf, H. S. (1965). Uniqueness theorems for periodic functions. *Proceedings of the American Mathematical Society*, 16, 109–114.

- Fogg, N. P. (2002). *Substitutions in dynamics, arithmetics and combinatorics*, volume 1794 de *Lecture Notes in Mathematics*. Edité par V. Berthé, S. Ferenczi, C. Mauduit et A. Siegel.
- Fraenkel, A. S. et Simpson, J. (1998). How many squares can a string contain? *Journal of Combinatorial Theory. Series A*, 82(1), 112–120.
- Gawrychowski, P., Kociumaka, T., Rytter, W. et Walen, T. (2015). Tight bound for the number of distinct palindromes in a tree. Dans *String Processing and Information Retrieval - 22nd International Symposium, SPIRE 2015*, 270–276.
- Glen, A. et Justin, J. (2009). Episturmian words: a survey. *Theoretical Informatics and Applications. Informatique Théorique et Applications*, 43(3), 403–442.
- Glen, A., Justin, J., Widmer, S. et Zamboni, L. Q. (2009). Palindromic richness. *European Journal of Combinatorics*, 30(2), 510–531.
- Gowers, T. (13 juillet 2012). What are dense Sidon subsets of  $\{1, 2, \dots, n\}$  like? Récupéré de [gowers.wordpress.com/2012/07/13/what-are-dense-sidon-subsets-of-1-2-n-like/](http://gowers.wordpress.com/2012/07/13/what-are-dense-sidon-subsets-of-1-2-n-like/)
- Hussini, S., Kari, L. et Konstantinidis, S. (2003). Coding properties of DNA languages. *Theoretical Computer Science*, 290(3), 1557–1579.
- Jeuring, J. (2013). *Finding palindromes: variants and algorithms*. Rapport technique UU-CS-2013-016, Department of Information and Computing Science, Utrecht University.
- Kari, L., Kitto, R. et Thierrin, G. (2002). Codes, involutions, and DNA encodings. Dans W. Brauer, H. Ehrig, J. Karhumäki, et A. Salomaa (dir.). *Formal and Natural Computing*, volume 2300 de *Lecture Notes in Computer Science*, 376–393.
- Kari, L. et Mahalingam, K. (2010). Watson-Crick palindromes in DNA computing. *Natural Computing*, 9(2), 297–316.
- Labbé, S. (2008). *Propriétés combinatoires des  $f$ -palindromes*. Université du Québec à Montréal, mémoire de maîtrise.
- Labbé, S. et Reutenauer, C. (2015). A  $d$ -dimensional Extension of Christoffel Words. *Discrete & Computational Geometry. An International Journal of Mathematics and Computer Science*, 54(1), 152–181.
- Lothaire, M. (1983). *Combinatorics on Words*, volume 17 de *Encyclopedia of Mathematics and its Applications*. Addison-Wesley Publishing Co.
- Lothaire, M. (2002). *Algebraic Combinatorics on Words*. Cambridge: Cambridge



University Press.

- Morse, M. et Hedlund, G. A. (1938). Symbolic dynamics. *American Journal of Mathematics*, 60, 815–866.
- Morse, M. et Hedlund, G. A. (1940). Symbolic dynamics II. Sturmian trajectories. *American Journal of Mathematics*, 62, 1–42.
- OEIS Foundation Inc. (2011). The on-line encyclopedia of integer sequences. Récupéré de [www.oeis.org](http://www.oeis.org)
- Pansiot, J.-J. (1984). Complexité des facteurs des mots infinis engendrés par morphismes itérés. In *Automata, languages and programming*, volume 172 de *Lecture Notes in Computer Science* 380–389.
- Pelantová, E. et Starosta, Š. (2013). Languages invariant under more symmetries: overlapping factors versus palindromic richness. *Discrete Mathematics*, 313(21), 2432–2445.
- Prouhet, E. (1851). Mémoire sur quelques relations entre les puissances des nombres. *Comptes rendus hebdomadaires des séances de l'Académie des sciences*, 33(225).
- Sidon, S. (1932). Ein Satz über trigonometrische Polynome und seine Anwendung in der Theorie der Fourier-Reihen. *Mathematische Annalen*, 106(1), 536–539.
- Starosta, Š. (2011). On theta-palindromic richness. *Theoretical Computer Science*, 412, 1111–1121.
- Thue, A. (1906). Über unendliche Zeichenreihen. *Norske Videnskabers Selskabs, Skrifter, I Mathematisch-Naturwissenschaftliche Klasse*, (7), 1–22.
- Thue, A. (1912). Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. *Norske Videnskabers Selskabs, Skrifter, I Mathematisch-Naturwissenschaftliche Klasse*, (1), 1–67.